

103  
**REVIEW OF INTERIM REPORT BY THE NATIONAL  
ACADEMY OF SCIENCES ON CENSUS REFORM**

---

Y 4.P 84/10:103-12

Review of the Interim Report by the...**ING**

**BEFORE THE  
SUBCOMMITTEE ON CENSUS, STATISTICS AND  
POSTAL PERSONNEL  
OF THE  
COMMITTEE ON  
POST OFFICE AND CIVIL SERVICE  
HOUSE OF REPRESENTATIVES**

**ONE HUNDRED THIRD CONGRESS**

**FIRST SESSION**

**MAY 27, 1993**

**Serial No. 103-12**

Printed for the use of the Committee on Post Office and Civil Service



APR 19 1994

U.S. GOVERNMENT PRINTING OFFICE

76-498

WASHINGTON : 1994

For sale by the U.S. Government Printing Office  
Superintendent of Documents, Congressional Sales Office, Washington, DC 20402

ISBN 0-16-043630-3



85  
**REVIEW OF INTERIM REPORT BY THE NATIONAL  
ACADEMY OF SCIENCES ON CENSUS REFORM**

---

Y 4. P 84/10: 103-12

Review of the Interim Report by the...**ING**

BEFORE THE  
SUBCOMMITTEE ON CENSUS, STATISTICS AND  
POSTAL PERSONNEL  
OF THE

COMMITTEE ON  
POST OFFICE AND CIVIL SERVICE  
HOUSE OF REPRESENTATIVES

ONE HUNDRED THIRD CONGRESS

FIRST SESSION

MAY 27, 1993

**Serial No. 103-12**

Printed for the use of the Committee on Post Office and Civil Service



APR 19 1994

U.S. GOVERNMENT PRINTING OFFICE

76-498

WASHINGTON : 1994

---

For sale by the U.S. Government Printing Office  
Superintendent of Documents, Congressional Sales Office, Washington, DC 20402  
ISBN 0-16-043630-3

## COMMITTEE ON POST OFFICE AND CIVIL SERVICE

WILLIAM L. CLAY, Missouri, *Chairman*

PATRICIA SCHROEDER, Colorado  
FRANK McCLOSKEY, Indiana  
GARY L. ACKERMAN, New York  
THOMAS C. SAWYER, Ohio  
PAUL E. KANJORSKI, Pennsylvania  
ELEANOR HOLMES NORTON, District of  
Columbia  
BARBARA-ROSE COLLINS, Michigan  
LESLIE L. BYRNE, Virginia  
MELVIN L. WATT, North Carolina  
ALBERT RUSSELL WYNN, Maryland  
GREG LAUGHLIN, Texas  
SANFORD D. BISHOP, Jr., Georgia  
SHERROD BROWN, Ohio  
ALCEE L. HASTINGS, Florida

JOHN T. MYERS, Indiana  
BENJAMIN A. GILMAN, New York  
DON YOUNG, Alaska  
DAN BURTON, Indiana  
CONSTANCE A. MORELLA, Maryland  
THOMAS J. RIDGE, Pennsylvania  
THOMAS E. PETRI, Wisconsin  
SHERWOOD L. BOEHLERT, New York

GAIL E. WEISS, *Staff Director*

ROBERT E. LOCKHART, *General Counsel*

DORIS MOORE-GLENN, *Deputy Staff Director*

JOSEPH A. FISHER, *Minority Staff Director*

---

## SUBCOMMITTEE ON CENSUS, STATISTICS AND POSTAL PERSONNEL

THOMAS C. SAWYER, Ohio, *Chairman*

FRANK McCLOSKEY, Indiana  
ALBERT RUSSELL WYNN, Maryland

THOMAS E. PETRI, Wisconsin  
THOMAS J. RIDGE, Pennsylvania

TERRIANN LOWENTHAL, *Subcommittee Staff Director*

## CONTENTS

---

	Page
Hearing held in Washington, DC, May 27, 1993.....	1
Statement of:	
Hunt, William M., Director, Federal Management Issues, General Government Division, U.S. General Accounting Office; accompanied by Bruce Johnson, Assistant Director, and Jack Kaufman, Senior Evaluator .....	72
Scarr, Harry A., Acting Director, Bureau of the Census; accompanied by Susan Miskura, Chief, 2000 Census Research and Development Staff .....	54
Schultze, Charles, Chair, Panel on Census Requirements in the Year 2000 and Beyond, Committee on National Statistics, National Academy of Sciences; accompanied by Barry Edmonston, staff director .....	3
Prepared statements, letters, supplemental materials, et cetera:	
Scarr, Harry A., Acting Director, Bureau of the Census:	
Prepared statement of .....	57
Response to written questions submitted by Congressman Sawyer .....	66
Schultze, Charles, Chair, Panel on Census Requirements in the Year 2000 and Beyond, Committee on National Statistics, National Academy of Sciences .....	10
Hunt, William M., Director, Federal Management Issues, General Government Division, U.S. General Accounting Office:	
Prepared statement of .....	75
Response to written questions submitted by Congressman Sawyer .....	83



# REVIEW OF INTERIM REPORT BY THE NATIONAL ACADEMY OF SCIENCES ON CENSUS REFORM

THURSDAY, MAY 27, 1993

HOUSE OF REPRESENTATIVES,  
SUBCOMMITTEE ON CENSUS, STATISTICS AND POSTAL  
PERSONNEL,  
COMMITTEE ON POST OFFICE AND CIVIL SERVICE,  
*Washington, DC.*

The subcommittee met, pursuant to call, at 1:03 p.m., in room 311, Cannon House Office Building, Hon. Thomas C. Sawyer (chairman of the subcommittee) presiding.

Members present: Representatives Sawyer and Petri.

Mr. SAWYER. Given the exigencies of this afternoon, we may as well get started because it is uncertain where we will wind up. It is a pleasure, nonetheless, to convene this hearing to hear the Interim Report of the National Academy of Sciences special panel on census requirements. We also want to talk a little bit about 2000 census planning.

In 1991, we asked the Academy to look at three issues that relate to the census. They included: assessing the goals of the census, a comprehensive review of census methods, and consideration of alternative ways to meet the Nation's data needs.

I don't think it comes as any surprise to anybody that the census, as the cornerstone of the national statistic system, is at a crossroads. Costs are rising, and by some measures, accuracy is diminishing. And perhaps most importantly, we seem to be failing to measure some of the things that count when they count the most.

The truth is that the Bureau deserves enormous credit for its role in rethinking the census. We have simply tried in this institution, this body, to broaden the discussion to include the Academy and GAO.

But it is not too much to say that this independent review was prompted by a measure of frustration with the outcome of 1990. Perhaps expectations were too high. Still, for twice as much money, the 1990 count was less accurate than the one before it.

It is not that the Bureau didn't try. It was an admirable effort. It is just that, and these are not my words, but others who are knowledgeable, traditional census methods alone simply do not count the country anymore. If it is true that the traditional design no longer meets the data needs of the Nation, it is true probably more than anything because the Nation itself is changing more rapidly than our traditional ability to measure it.

So that brings us to the three basic questions that are in the mind of the Congress as we plan for the next census. Can we improve the accuracy of the population counts? Can we contain the costs and continue to reach primary goals? And are there alternative ways to improve timeliness, and as a result, fundamental accuracy and usefulness of the data?

Let me just talk about each one briefly. First, of course, the foremost concern is with the accuracy of the population numbers. It is the primary goal of the census. It is fundamental to understand that the count of the population really is at the heart of our Democratic system of governance. We are concerned about overall accuracy, but we are perhaps every bit as concerned about the relative accuracy among population subgroups.

The second overarching concern is cost. My colleagues all across the Congress, regardless of their view or the measure of sophistication they bring to these questions, clearly want to have the census that they pay for. They want to make sure that it counts the country. They are willing to provide a reasonable amount of money to get the job done, but they are not willing to pay more and perceive that they are getting less.

Costs are escalating rapidly, but the disproportionate undercounts of minorities also increased. And perhaps most disturbing of all is the fact that, in terms of trying to direct policy and understand how we are changing as a Nation, the demographic data for program purposes is becoming less meaningful when we first see it, just by virtue of the passage of time. Because the information is old when we get it, from a legislator's viewpoint, we are getting less bang for the buck.

Those two concerns, accuracy and cost, really lead to the final point, and that is we need to explore alternative ways to measure a wide range of data that we traditionally count in the census.

There is a growing perception that the census has become unwieldy and that the level of content collected on the forms may be diminishing the accuracy of the population counts. The Bureau's research supports that fear. Greater content yields lower mail response. And when mail response goes down, accuracy suffers measurably. In order to prop that up, the costs associated with door-to-door follow-up are enormous.

Finally, the relationship between content and coverage isn't the only reason to consider the redistribution of the data burden throughout the decade. All of us need data that is more accurate. It is not enough just that the data be precise. In a period of rapid change, if the numbers aren't timely, they are not accurate. And what looks most precise is often the most fragile information. It becomes inaccurate and too often irrelevant most quickly by virtue of that time and movement.

When change is the, or one of the dominant characteristics that we are trying to measure in the course of a decade, and a census occurs only once every decade, then it seems to me we are failing to capture what may be most important to measure.

In any event, I look forward to hearing the thoughts of the Academy today and the Bureau and GAO on each of those important issues.



Our first panelist today is Dr. Charles Schultze who chairs the Panel on Census Requirements in the Year 2000 and Beyond, the Committee on National Statistics at the National Academy of Sciences. It is a pleasure to have you here today.

**STATEMENT OF CHARLES SCHULTZE, CHAIR, PANEL ON CENSUS REQUIREMENTS IN THE YEAR 2000 AND BEYOND, COMMITTEE ON NATIONAL STATISTICS, NATIONAL ACADEMY OF SCIENCES; ACCOMPANIED BY BARRY EDMONSTON, STAFF DIRECTOR**

Mr. SCHULTZE. Thank you, Mr. Chairman, and also thank you for rescheduling the time to fit my schedule. I appreciate that.

Mr. SAWYER. If we could pause for just a moment. Tom, do you have an opening statement you wanted to offer or any thoughts or observations?

Mr. PETRI. No, except that I am glad I got here to hear Mr. Schultze because I have been one of his fans for a very long time.

Mr. SAWYER. We both share that. If you would like to add anything for the record, that's fine.

Dr. Schultze.

Mr. SCHULTZE. Mr. Chairman, you have our report. That report has a summary. Let's consider that my formal testimony. And what I want to do, I am not even going to summarize as such, but call your—

Mr. SAWYER. Without objection, we will just include the entire body of your report in the record for today.

Mr. SCHULTZE. Thank you, Mr. Chairman. What I am going to do is call attention to some of the highlights in that report, rather than give you a comprehensive summary. And so if I don't mention something, it isn't because we didn't cover the subject. It may still be there.

This is, as you know, our interim report. We are shooting—we are instructed to shoot for a November 1994 final report. This report, therefore, includes only those recommendations with some time urgency. We deliberately made a choice not to tie ourselves in knots worrying about specific recommendations at this stage and, therefore, you don't have many. You do have three because they were time urgent.

Let me try to give you some idea what the report does, also, of how we are going about our task and some of our preliminary views, but let me stress preliminary. We may even be so bold as to change our minds if evidence, further evidence, and our studies warrant it.

Let me also stress, as I did the first time I testified before your committee, that we are taking a long-term look at the Census in the Year 2000 and Beyond and we don't really believe that our principal responsibility is to oversee or second guess the planning process for the year 2000 census, except to the extent it would lock us in long term or miss some sort of testing that would be helpful for long-term changes.

Let me turn first to the problems of cost and quality in the current census process, covering two of your points. You are of course aware of the rising cost of the Census, from about \$11 per housing unit in 1970 to \$25 in 1990, and under certain mildly pessimistic

conditions, perhaps \$30 by the year 2000, in inflation adjusted dollars. So it would triple over this 30-year period if things go as it now looks and if nothing is done.

And as you yourself just noted, simultaneously—in the grossest sense of the word—the quality of the results is deteriorating. The overall undercount was larger in 1990 than in 1980 and so was the differential undercount. We think there are two, at least two, not to say the only two, important driving forces, especially over the last 10 years.

The response rate fell and fell pretty significantly from 75 percent to 65 percent. Public cooperation, then—not merely here but in all sorts of other survey type and public civic type activities—has unfortunately declined. And simultaneously, the combination of court rulings, the modifications and amendments to the Voting Rights Act, have increased the demand for accurate data by race, ethnicity and age, and not only for large geographic groupings, but for small areas. Moreover, the response rate is worse precisely where the problems of the undercount are most serious.

These two developments, the way the census has responded to them, and perhaps even more important, the way they are perceived by the public, have interacted to raise costs substantially without improving quality. The panel is still studying this whole set of issues, but we would like to make three preliminary observations.

First, the Bureau of the Census sought to deal with this very real set of problems—the falling response rate and increased demand for fine-grained detailed data—by heavily labor intensive and highly expensive follow-up techniques and coverage improvements, seeking to secure a physical count of everybody.

An alternative is to use sampling techniques as a way to follow up on the mail questionnaire or in other initial contacts, and then to survey very intensively some areas and use models and sampling to impute the results to the totals. It is quite likely, but clearly needs more study that, with no sacrifice in overall accuracy, this approach might yield substantially reduced costs with, as I say, perhaps no and very possibly even improved accuracy.

Substituting sampling for physical enumeration is a major area we have concentrated on and want to look at further and which the Census itself will explore during its tests. Legal studies that we have asked to be done suggest that as long as there is a bona fide attempt through, for example, mailed questionnaires to reach everybody, this would satisfy constitutional requirements for enumeration. Although our main charge is not to become constitutional lawyers, that is the way we now read it.

Mr. SAWYER. Are you satisfied being constitutional scientists?

Mr. SCHULTZE. We may actually need some more, you know, outside opinion. But that is the way it now looks.

Mr. SAWYER. I understand.

Mr. SCHULTZE. Not only that, this approach would allow the Census to concentrate its main efforts where the problems are, in areas of particularly heavy differential undercount. And so in this interim report, one of our recommendations is that serious consideration be given by the Census Bureau to investigating sampling for nonresponse follow-up in the 1995 census test.

We recommend testing to provide information on the costs, the effects on small area data, and whatever statistical problems may emerge. And it is my understanding, reading the preliminary census material on this, that they are proceeding along those lines.

Second major observation: Much attention has been given to the possibility of reducing content as a means of lowering census costs. In the first place, changing content is not the cause of the rise in costs. For example, the long form has remained roughly the same for some period of time and the sampling is lower. Yet, costs are marching up.

On the evidence to date, it appears to us that modest changes in content would have little impact on costs. The marginal cost of taking one, two, three, four, five questions off or on, is not very great. And indeed, even if you look at the whole long form, the cost is apparently something like \$200 million for having a long form questionnaire. If you dropped it, you would save something like \$200 million.

However, the panel is also aware of the possibility that as you yourself mentioned, Mr. Chairman, very large differences in complexity and the length of the census could affect the response rate and, therefore, costs. It is very hard to get evidence on that. You have a little evidence. It turns out that the response rate for the short form was 66 percent and that ones including the long form, 60. Well, that is not massive, but it is not zero. We will get more information from some of the current census tests by contrasting response rates with radically different forms and we will continue to pursue this.

We also think it is important to distinguish content from the friendliness of the form. And it is important not to mix up the two. You can have perhaps fairly substantial content with more or less friendly ways of looking at it. Perhaps we could ask the IRS to help us with that.

I hope the transcript will show that that was said with a smile.

We have made some preliminary investigations and plan in the future to examine very carefully the uses of data collected on the long form, both in itself and as background support for intercensal surveys. We want to look at the potentials and the costs of various alternative mixes of the current census approach, the use of administrative records, and larger intercensal survey.

Let me simply note that from the evidence that we have collected, the cost of the existing annual surveys, which are absolutely critical to get us data on income and consumer expenditures, would be significantly greater without a census sampling frame. There is an interaction here. So we are, let us say, open-mindedly skeptical about achieving large results from changing content but realize we have to pursue this in various ways.

Third point: There is a mistaken belief that to be useful data must be highly accurate at the block level. That is impossible in any event. You are going to get fairly sizeable errors in individual block data when you are breaking it down by race, ethnicity, age, and the like.

What is relevant is the accuracy of the data when it is aggregated to some higher level, for whatever purpose you want. For example, the accuracy of population counts and ethnicity data in a con-

gressional district is very important. But it is important to distinguish the two. If bias isn't present, offsetting errors will give a high degree of accuracy at, say, a congressional level area, even though, block by block, it is not terribly accurate.

We believe it would be very helpful to the public in an indirect sort of way and to the users of census data, including the courts, to get an analysis of errors and biases at the block level so that, as I said, users of the census data, including the courts, can understand the nature of the block level data they are starting with. Perhaps it might even relieve some of the pressure for unattainable accuracy at very, very small levels.

And so, recommendation number two is that the Bureau of the Census analyze the 1990 post-enumeration survey data, which I gather now is technically feasible to do, to produce estimates of gross errors, that is pluses and minuses, at the block level in the census. And that is required to examine census requirements for the accuracy of small area data in the future. It may be the fear of public excoriation for being inaccurate at these very low levels that may have driven the costs to some extent, by what I might call excessive follow-up.

In addition to the problem of cost, the panel has spent a good bit of time examining and analyzing the constitutional, legal and other considerations of public policy that determine national data requirements. After all, that is the charge to our panel, is a requirements panel. Such an analysis is essential for considering the feasibility and desirability of alternative census designs.

Let me start with the most fundamental constitutional requirement for an enumeration, every 10 years for reapportionment purposes. You have to count not only the population, but you have to assign every person to an address. Mere accuracy in the count isn't enough.

And third, not only will you have to have an overall count assigned to an address, but we have to have it—at a minimum to carry out the laws of the Congress—by race, ethnicity, age, and in small area groups assigned by residence.

It is important to remember that, because this begins to shape what can and can't be done by way of alternative designs. It is our tentative judgment you can't have a sample census if you look at it that way, purely a sample census. But you can obviously sample as part of an overall census.

Second, this is of major importance in considering the use of administrative records for the basic census itself. IRS, Social Security, many State and local records have marvelous amounts of information, but very often the addresses are of course no good. Social Security data might be 30 years old. And age, ethnicity, is just about on none, and there are very delicate problems about putting it on.

Europeans, some European countries have systems that do this. The Danes, the Finns, the Dutch, have a central register, continuously updated in which individuals are assigned a residence continuously. And that does mean that you pull together the two sets of information so that you can meet the geographical assignment requirements of the census and then begin to link other things into it.



But what would people in the United States think about this? This is not merely a technical matter, it is a matter of a whole culture. But it needs to be pointed out that very often people, what is the word I want, blithely assume that all this marvelous information and administrative records could be used but it is a major problem, which is not to say we are not sympathetic to doing it. Not to say it can't be done, but it involves all sorts of things. And there are trade-offs.

So we first agree with our companion panel in its letter to the Census that administrative records really are not a basis for the year 2000 census, but we would like to pursue it and plan to pursue it heavily, what some of the options are for the longer term.

And, therefore, we recommend in our recommendation number three, that the Bureau of the Census should initiate a separate program of research and uses of administrative records, not for the year 2000 itself, focusing primarily on the 2010 census, and on current intercensal estimates program; to undertake a planning study to develop detailed design options for 2010 administrative record census, because only if you do it in detail, you can begin to realize what the problems are; and to seek the cooperation of Federal agencies that maintain key administrative record systems in undertaking experimental mini censuses and related projects based on administrative records and give priority where it is feasible in the year 2000 census to supplemental use of those records, even though they can't be—it is our judgment, you can't really build them in as a major part.

Now, in addition to the enumeration and the voting rights and related data, there are a whole batch of other needs of course in the census. The Congress itself has almost had an explosion in the use of census or census—based data for apportioning funds. If you look at a compilation of the relevant statutes, far less look at what they ask for, just look at a compilation of the statutes, it is an impressively sized list.

The Federal statistical system itself relies heavily on census data, in many cases simply for establishing some kind of a frame to do surveys at reasonable cost. You would have to have much bigger surveys if you couldn't relate them to the census framework. And of course State and local governments use them for all sorts of purposes. And finally, private uses are sharply increasing, particularly the use of census data by block as a building unit for private compilations of data.

For all of these uses, timeliness, as again you said yourself, is very important. And it is probably more important in the United States than almost any other society precisely because we are more mobile. Ten minutes after the census is done, it has started to become wrong. Originally good data 8 to 10 years out of date are no longer good data.

As a general proposition, eliminating some types of information from the census and substituting periodic, annual, biannual surveys, cannot be done except at much higher costs over the whole 10-year period. They will, on the other hand, provide much more timely data, but also the system then becomes more expensive, especially for small area data.

So you do have a trade-off here between costs and timely data. And at the moment, it is not for me to say which way you ought to push, but it is one of those unfortunate choices where you can't have your cake and eat it. And timeliness may turn out to require more money.

We want to investigate, however, the potential uses of administrative records, Federal, State and local, for intercensal estimates of various kinds, and the possibilities they are perhaps more feasible in a shorter period of time than actually substituting administrative records for the census itself.

Tying in the census records and using the administrative data, there are all sorts of possibilities, some of which we need to explore ourselves in the panel and at least block out some directions for further exploration in the future, because timeliness is a critical element. And the cost of doing it by the survey route is very expensive.

Now, all of this, when you consider administrative records, raises a number of issues that have to be resolved. The panel will consider and, if necessary, provide the Congress with options and, where appropriate, point out necessary legislative changes in several areas. First, procedures, laws and regulations may need to be changed or rewritten to allow Federal agencies to share administrative records with each other and with the Census for statistical purposes subject to appropriate safeguards for confidentiality. And, of course, the key word in there is appropriate and people will argue over what is appropriate.

Second, there is the whole broad question of linking Federal, State, and local administrative records to a geographic database, and that has to be considered if these data are to be useful in many cases. For example, through the Census Bureau's address system and TIGER Geo Coding, data can be provided classified and cross-tabulated by small geographic areas, but again it will be absolutely necessary to develop procedures and safeguards to allow the census geographic database to be linked to State and local and Federal administrative data. And, again, the panel will at least try to spell out the options and the trade-offs.

The report briefly discusses its future work load, what we plan to do in the period between now and when we are due to give you a report. We are going to continue to pursue the issues I have already discussed with two objectives: in some cases to provide Congress with specific recommendations, and in other cases, to point out areas where basic political choices have to be made by the Congress, spelling out some specific alternatives and trade-offs.

For example, the use of administrative records and the associated problems of confidentiality, and privacy. We are going to emphasize the very large issue of sampling for follow-up. There is a major cost-saving potential.

Second, we want to do a careful analysis of the various factors that have in fact been driving costs up. We think we have got the basic ideas down but there is still more work to be done.

Third, we want to examine further the long term, that is post-2000 potentialities of using administrative records as a substitute for the census, fully or partially. And we will try to spell out the

necessary changes in administrative records that would be required and the associated issues of confidentiality.

We will examine specific alternative uses for administrative records for more timely intercensal estimates, including the various implications and consequences of tying together administrative records with a geographical database as a means of providing Congress, the executive branch and State and local governments with more timely data for policy planning, fund allocation and operations.

In addition, and finally, Mr. Chairman, are two major areas that we have not addressed in any depth today to which we will turn attention, and that is the issues that arise in connection with providing, as has been the case in recent censuses, of increasingly detailed ethnic classifications in the census.

Experience in the United States and Canada suggests that answers to questions of ethnicity vary substantially depending on the context. If people are encouraged to do so, they are more likely to specify their ethnic background than if it is simply left as an option.

It is more likely that they will specify it in censuses than in administrative records probably because the power of the police state is still hanging over one much more than the other. This is a policy matter. To what extent should it be encouraged or discouraged, or should public policy try to stay neutral. You can't actually stay neutral. So the whole question depends on the context and we are can't assume it is something that is God given out of nature; people will answer the ethnicity question quite differently depending on the overall context in which they are doing it.

We want also to explore the question of what the ethnic data means, given the increasing frequency of marriage across ethnic and racial lines.

We also want to look at the question of whether and to what extent increases in the number of classifications and the effort to provide accurate ethnic data by small areas in all parts of the country has played a role in raising costs and to what extent would it be feasible and desirable to tailor some of the questions to particular parts of the country. As I said, we have not addressed these issues, but we plan to.

We haven't really decided this yet, but I think it is almost surely necessary that we spend some time taking a look at the extent to which, if at all, the changing structure of the household and families in the United States has any implications for longer range Census planning. In general, yes, it surely does, but in specific we have not yet done anything on it.

Finally, Mr. Chairman, let me conclude by saying I think in all of this, we have to recognize that the census and the traditions that are incorporated in it are an asset to the country and a very important asset. It does provide us with the background for an amazing statistical system. It needs reform, maybe some radical reform, but we have to be very careful and the panel takes this very seriously, to do this very responsibly, because the last thing we want is to destroy an existing asset for pie in the sky reform.

Thank you, Mr. Chairman.

[The prepared statement of Mr. Schultze follows:]

PREPARED STATEMENT OF CHARLES SCHULTZE, CHAIR, PANEL ON CENSUS REQUIREMENTS IN THE YEAR 2000 AND BEYOND, COMMITTEE ON NATIONAL STATISTICS, NATIONAL ACADEMY OF SCIENCES

PREFACE

The Panel on Census Requirements in the Year 2000 and Beyond is conducting a study to assess the needs for data currently collected in the decennial census. In the first part of the panel's work, summarized in this interim report, the panel has been assessing census designs for accurate, cost-effective data collection for achieving a count of the population, as required by the Constitution. In the second part of the panel's work, to be presented in its final report in late 1994, the panel will review information on the costs of alternative methods, assess how timely small-area data can be provided over a decade, investigate the degree to which a continuing (or new) need is anticipated with respect to the types of data collected in the census, and examine census designs for achieving an accurate population count, taking cost into account.

The Panel on Census Requirements in the Year 2000 and Beyond began its work in the summer of 1992 and held several panel meetings during its initial deliberations. During its first year of activities, the panel has discussed the major requirements for the census, debated the criteria for a census, and considered the tradeoffs in trying to achieve a census that fulfills the basic needs for reapportionment and redistricting, for the federal statistical system, for users of small-area data, and for analyses of small population groups. The panel met for a considerable time in working groups.

One working group is considering small-area data and small population groups; it is chaired by Michel Lettre. This group has met with state and local data users, representatives of national ethnic advocacy groups, and individuals knowledgeable about the use of census data by corporations and nonprofit enterprises.

Another working group is considering the federal statistical system; it is chaired by Janet Norwood. This group has discussed census data needs with officials from several large federal statistical agencies and with officials from the U.S. Office of Management and Budget.

A final working group is considering issues related to reapportionment and redistricting; it is chaired by Stephen Fienberg. This group has benefited from the recent report on legal issues for the 2000 census, prepared by the Congressional Research Service, and from comments on the Congressional Research Service report from legal scholars. In addition, the staff has interviewed individuals who use census data for redistricting. The group is currently soliciting advice from scholars familiar with the Voting Rights Act.

The panel wishes to thank a number of people who contributed to the preparation of this report. Several Bureau of the Census staff gave helpful assistance. Barbara Everitt Bryant, who was director of the Bureau when the panel began its work, and Harry Scarr, who is currently acting director, provided an overview of census planning and offered valuable briefings on current census activities. Other knowledgeable staff who helped the panel with advice and information include Robert Tortora, Susan Miskura, James Dinwiddie, Jay Keller, Joe Knott, and Robert Marx.

At the beginning of the panel's activities, the panel was pleased to be briefed by U.S. Representatives Harold Rogers and Tom Sawyer on their views of the decennial census and their thoughts on criteria for future censuses. We appreciate assistance received from a number of congressional staff, including Kevin Fromer, TerriAnn Lowenthal, Shelly Wilkie Martinez, and David McMillen.

Staff from other federal agencies provided briefings and information to the panel. We acknowledge the assistance of Bruce Johnson, Chris Mihm, and Jack Kaufman from the U.S. General Accounting Office; Jennifer Williams and Margaret Mikyung Lee from the Congressional Research Service; Fritz Scheuren, Peter Sailer, and Ellen Yau from the Internal Revenue Service; Wesley Schaible and William Barron from the Bureau of Labor Statistics; Maria Gonzalez and Katherine Wallman from the U.S. Office of Management and Budget; and Monroe Sirken from the National Center for Health Statistics.

The panel met with representatives of various groups and other individuals interested in the provision of census data. We thank these individuals for taking the time to share their opinions with the panel: Lorraine Amico, National Governors' Association; Hamilton Brown, National Association of Towns and Townships; Donald F. Cooke, Geographic Data Technology; David Crowe, National Association of Home Builders; Norman Deweaver, Indian and Native American Employment and Training Coalition; Jon Felde, National Conference of State Legislators; Charles Kamasaki, La Raza; Monica Kuumba, National Urban League, Inc.; Sharon Law-



rence, National Association of Counties; William O'Hare, University of Louisville; Thomas Palmerlee, Urban and Regional Information Systems Association; Martha Riche, Population Reference Bureau; and Lance Simmens, U.S. Conference of Mayors.

The panel also thanks Larry D. Barnett of Widener University and Samuel Issacharoff of the University of Texas, Austin, for providing comments on the constitutional and statutory mandate for the census and the legal requirements of the census.

Several staff members of the Committee on National Statistics offered useful guidance and provided information, especially Miron Straf, Edwin Goldfield, Duane Steffey, and Meyer Zitter. The panel also appreciates the effort of Eugenia Grohman, associate director for reports of the Commission on Behavioral and Social Sciences and Education, for her skillful editorial work, which greatly improved the presentation and argument of this report.

The panel owes thanks to its own staff. Michele Conrad handled the administrative management of the panel with consummate skill, dealt with the logistics of the panel's meetings, and competently handled the preparation of the various drafts of this report. Constance Citro worked closely with the panel's working group on reapportionment and redistricting and played a leading role in the presentation of that group's work in this report. Barry Edmonston, who had primary responsibility for the overall preparation of this report, is to be thanked for his fine efforts. The report would not have been possible without the dedicated effort of the staff.

Finally, I would like to thank the panel members for their contribution of time and advice. Several members provided background documents and other members prepared drafts of sections of this report. I am delighted that the panel has had vigorous debate on the issues of census requirements. I look forward to continued work with a fine panel.

CHARLES L. SCHULTZE,

*Chair, Panel on Census Requirements in the Year 2000 and Beyond.*

#### EXECUTIVE SUMMARY

At the request of the Congress, and with sponsorship from the Bureau of the Census in the U.S. Department of Commerce, the Panel on Census Requirements in the Year 2000 and Beyond is conducting a study on the needs for data that are currently collected in the census. It is identifying the data for which the census is required or is the most effective means of collection and studying cost-effective methods for carrying out the constitutional requirement for a count of the population and meeting other data needs.

This interim report presents a discussion of the panel's findings to date. We include in this report three recommendations on topics for which there is a time urgency. The final report, to be completed in late 1994, will include recommendations on census planning and census designs that the panel believes deserve consideration for the future.

The key foci of the panel's initial work have been the major requirements for the census, the criteria for a census, and the tradeoffs in trying to achieve a census that fulfills the basic needs for reapportionment and redistricting, for the federal statistical system, for users of small-area data, and for analysis of small population groups. It has become clear to the panel that the census serves a large number of purposes, including constitutional and statutory data needs for reapportionment and redistricting, mandatory data needed for federal programs and federal funding allocations, and data for many other uses by public and private groups.

The panel's deliberations have identified four major issues regarding the decennial census: rapidly escalating costs, a widening difference in the undercount between minorities and the white population between 1980 and 1990, continued pressure for accurate small-area data from many census data users, and increasing demand for more timely data. The preliminary assessment of the panel is that there is no panacea to address simultaneously these four issues. The report discusses some of the proposed remedies: a shorter census form, continued effort at coverage improvement, and possible alternative census designs.

The panel began its work with a study of the legal requirements for the census. The major conclusion reached by the panel is that a census must include an attempt at a basic enumeration to meet constitutional requirements. Such a complete enumeration of the population once a decade, would rule out such census designs as a rolling census or a sample census from consideration.

The report discusses several technical issues, including sampling for nonresponse follow-up, the accuracy of block data, the timeliness of data, and small-area data and a geographic database. The report also covers some possible new census designs,

with particular emphasis on expanded uses of administrative records. They will be considered in detail in the final report.

Finally, the report describes several key issues in census requirements that will be in the final report. For example, the census needs to provide an accurate count of racial and ethnic groups for small geographic areas. This is an important issue and the panel plans further study for its final report. The interim report includes some of the panel's deliberations on the use of administrative records for the census; the panel will continue to study this topic and will make recommendations in the final report.

One of the main costs in census operations is dealing with nonrespondents to the mail questionnaire. With declining mail response rates, it has become very expensive to send out enumerators to follow up with nonrespondents. Moreover, some of the "nonrespondents" turn out to be vacant housing units. One potentially cost-effective option for future censuses is the use of sampling for nonresponse follow-up. If sampling is used, a selected sample of nonrespondents would be contacted and then information would be inferred for all addresses from which a mail questionnaire was not returned.

The panel is concerned that sufficient information does not exist for evaluating the potential usefulness of sampling for the follow-up of nonrespondents to the census mail questionnaire. It is important to examine this methodology in the 1995 census tests. Only such testing can provide the requisite data needed for serious consideration of this technique for the 2000 census. Without such testing, there would be no information for discussion of its possible inclusion in the 2000 census design.

Recommendation 1. The panel recommends that serious consideration be given to investigating sampling for nonresponse follow-up in the 1995 census tests. The panel recommends testing to provide information on the costs, effects on small-area data, and statistical problems.

Census data users are increasingly relying on small-area data, including census block data. The census block data are the fundamental units for constructing the special areas required by legislative redistricting analysts, state and local planners, and academic researchers. In transportation planning, for example, local planners designate special transportation areas in their cities or counties and then aggregate block data for these small areas.

The accuracy of small-area data is an important issue for the evaluation of census data and for planning the design of future censuses. But at present there is a lack of information about the nature of errors for small-area data. Such information would provide the basis for discussion of what levels of accuracy for small-area data are needed in future censuses and how different census designs might affect the current levels of accuracy.

Recommendation 2. The panel recommends that the Bureau of the Census analyze 1990 Post-Enumeration Survey data to produce estimates of gross errors at the block level in the census. This information is required for examining census requirements for the accuracy of small-area data in the future.

Administrative records offer considerable potentials for improving census activities. Although administrative records were used to a limited extent in the 1990 census for coverage improvement, use of administrative records could be expanded for coverage and content improvement and for intercensal estimates.

Some European countries have discontinued direct enumeration and now rely completely on administrative records for a population census. The panel does not believe that administrative records could be used in the next decade to meet even relaxed demands for census data for small geographic areas. It is important, however, to look beyond the 2000 census to ask what information would be needed in order to consider the relative merits of using administrative records for the census. To do so, additional research must be conducted.

The panel concurs with and endorses recommendations made in 1992 by the Panel to Evaluate Alternative Census Methods regarding examination of the use of administrative records for the census.

Recommendation 3. The panel recommends that the Bureau of the Census should (a) initiate a separate program of research on uses of administrative records, not directly related to the 2000 census, focusing primarily on the 2010 census and on current estimates programs; (b) undertake a planning study to develop detailed design options for a 2010 administrative records census; (c) seek the cooperation of federal agencies that maintain key administrative record systems in undertaking experimental minicensuses and related projects based on administrative records; and (d) give priority to some use of administrative records in the 2000 census for those purposes for which such usage is feasible, such as coverage and content improvement and coverage evaluation.

## BACKGROUND: CONTEXT OF THE STUDY

The decennial census serves a wealth of important needs, including constitutional and legal requirements for reapportionment and redistricting of congressional seats, allocation of funds, and analysis of small areas and small population groups by all levels of government and private profit and nonprofit organizations. There was a big increase in census costs between 1970 and 1990 (in constant dollars), but the 1990 census provided coverage of the population that was no better—and possibly worse—than the 1980 census. The rising census costs and declining public cooperation in returning the census mail questionnaires occurred in the context of continued litigation over the accuracy of the census and widespread debate about adjusting the population counts. The problems encountered in taking the 1990 census, the perception of reduced accuracy, and the high and increasing costs led many Senators and Representatives in Congress to question what should be the fundamental approach to the census.

### *Charge to the Panel*

The Decennial Census Improvement Act of 1991 (P.L. 102-135) directs the National Academy of Sciences to conduct a study on the census in the year 2000 and beyond. The act requests a study of the means for achieving the most accurate population count possible and of how to collect other population and housing data. More specifically, the study is to look at improvements in census enumeration methods that collect information directly from respondents, as well as to investigate alternative methods for collecting data for a basic population count—including the use of administrative records, sample censuses, and cumulative or rolling censuses. The act asks for the study to consider the appropriateness of sampling methods for the acquisition of population data, including a review of the accuracy of data for geographic areas. Finally, the study is to examine the continuing need for the types of data that are collected in the census, whether there are more effective ways of collecting those data, and whether alternative sources or methodologies could provide more timely information.

The Committee on National Statistics of the National Academy of Sciences has responded to the congressional request by forming the Panel on Census Requirements in the Year 2000 and Beyond. This panel is evaluating the needs for data currently collected in the census and identifying for what data the census is either required or is the most effective means to collect the data; it will recommend accurate and cost-effective means for achieving through the census the constitutional requirement for a count of the population. The panel also will investigate ways for meeting other data needs, through the census or other means. The panel began work in 1992 and will deliver its final report in November 1994.

This panel is one of two census-related activities being carried out at the National Academy of Sciences. The other study, being conducted by a separate Committee on National Statistics Panel to Evaluate Alternative Census Methods, will provide scientific and technical evaluations of methods for taking the census in 2000 and subsequent years. The methods panel is sponsored by the Bureau of the Census; it will conduct its study through March 1994.

Both panels will recommend methods to achieve an accurate population count. The Panel to Evaluate Alternative Census Methods will review proposed methods and the Bureau of the Census's approach, provide technical evaluation, and suggest needed research in preparation for possible use in the year 2000 or later censuses. The Panel on Census Requirements in the Year 2000 and Beyond will recommend to the Congress methods for conducting future censuses and methods for meeting needs for data currently collected in the census, together with an analysis of the advantages and disadvantages of the methods. The two panels are taking independent approaches in their work.

### *Departmental Planning for the 2000 Census*

In addition to the two Committee on National Statistics studies of the decennial census, the U.S. Department of Commerce and the Bureau of the Census are exploring and developing data requirements, methodology, and operations for the decennial census in the year 2000. A task force for this purpose has been established under the Under Secretary for Economic Affairs in the Department of Commerce. The task force is considering 14 alternative methods for the collection of demographic, economic, and housing data. Some of these methods would be variants of the 1990 census design, some would use a minimal census with additional surveys, some would have an expanded role for administrative records, and some would involve a rolling census. The task force is composed of a technical committee, under the Bureau of the Census's Associate Director for Statistical Design, Methodology, and

Standards; a policy committee temporarily now under the acting director of the Bureau of the Census; and an advisory committee of representatives from a number of organizations.

The technical committee is primarily concerned with the evaluation of the statistical, operational, and methodological aspects of census-related activities, including alternative decennial census designs. Members of the technical committee are mainly from the Bureau of the Census, although it has also included representatives from the National Institute of Standards and Technology, the Internal Revenue Service, the Bureau of Labor Statistics, and the Department of Commerce.

The policy committee focuses on issues of the decennial census for the federal statistical system. This committee is examining policy implications of various proposals for decennial census designs, including cost, statutory and constitutional requirements, use of statistical sampling, new technologies, the relationship of the census to large federal surveys, and the implications of diversifying data collection procedures. Its members include representatives of the major users of census information, including the Departments of Defense, Education, Energy, Health and Human Services, Housing and Urban Development, Justice, Labor, and Transportation. The policy committee also includes a representative from the Office of Management and Budget.

The advisory committee seeks to ensure broad public participation in the discussion of the nature of the 2000 census. The advisory committee includes representatives from 25 organizations, representing private-sector data users, professional associations, minority groups, and state and local governments, along with executive and legislative branch officials as *ex officio* members.

Because the Congress and the Bureau of the Census contemplate the possibility of major changes in the conduct of the 2000 census, it is critically important to begin the planning process early. The schedule for census planning is important because many consequential decisions for the 2000 census are needed soon. By September 1993 a small number of census designs will be selected for testing in 1995. A final selection of the 2000 census design will be made by December 1995.

#### SCOPE OF INTERIM REPORT AND PLANS FOR FINAL REPORT

The Panel on Census Requirements in the Year 2000 and Beyond is in the midst of its investigations; it has not yet reached a point at which it is ready to make recommendations about the design, content, or operations of the census or alternative data collecting processes for the year 2000 or later.

This report outlines the areas of investigation the panel has undertaken and reviews, in a preliminary way, some of the important facts, issues, and analysis that have emerged to date from those investigations and from our deliberations. The panel's work to date covers four general areas:

- 1) understanding the nature, magnitude, and consequences—for both costs and quality—of the problems confronted by Congress and the Bureau of the Census in meeting increased demands for accurate and geographically fine-grained demographic and other data under conditions of deteriorating public cooperation with the census process;

- (2) studying the constitutional and legal requirements that, on one hand, determine important aspects of census data requirements and, on the other, eliminate some of the more radical alternatives to the traditional census;

- (3) investigating and discussing the costs, benefits, and legal implications of major alternatives and combinations of alternatives to the traditional census approach; and

- (4) identifying some of the difficult tradeoffs among competing objectives that will have to be made in choosing among alternative future paths for the collection of census-type data. The interim report contains three recommendations that have a time urgency to them.

While this interim report summarizes our work to date, the reader should be aware that at many points it represents preliminary thinking about a number of issues on which our analysis or our deliberations are incomplete. Additional facts and analysis and further deliberation by the panel could modify, especially in nuance or emphasis, some of what is reported here. In three instances, however, the panel's work has led to recommendations that are pertinent to decisions that will have to be made by the Bureau of the Census in the relatively near future about research, experimentation, and planning that are being undertaken for the 2000 census. Because of their time urgency, those recommendations are included here.

This report also includes two appendices: one summarizes the panel's work on the political and legal history of issues relating to census data needs for reapportion-



ment and redistricting; the other describes the development of new sources of information for small geographic areas.

The panel's final report will be completed in late 1994. It will address a number of topics concerning the requirements for the census, some of which are discussed without recommendations in this interim report. The request from Congress for a study of census requirements reflects the desire for a back-to-basics, zero-based consideration of the census that begins with no preconceived notions about what data are collected or how. The panel has responded to this request by considering, first, the constitutional requirements for a decennial census for the purpose of apportionment. The panel has also considered the major statutory requirements for the use of currently collected census data for congressional redistricting, including the requirements of the Voting Rights Act. The work of the panel on this topic is mandated by the congressional directive (The Decennial Census Improvement Act of 1991). The panel is also directed to recommend the most effective cost-effective and accurate methods of fulfilling the constitutional and mandatory requirements for a decennial census. The panel will report on this examination in its final report.

Future local, state, and national policies may require data that are not now collected in a usable way. For these data requirements, consideration must be given to the types of data and the level of geographic detail needed, as well as to current and potential sources of such data. The panel will give consideration to some important data requirements for such emerging policy issues as changing family structure, migration, transportation, and the labor force.

In looking to the future, the Bureau of the Census has encouraged consideration of a number of possibilities for the collection of demographic and housing data outside the traditional decennial census. Most of the proposed methods will not stand alone to provide small-area data, but must be used in combination with an intercensal estimation, a census program, or both. The panel is pursuing a study of new methodologies for taking a census and will include its findings in the final report.

Increasing demands for more detailed and current small-area estimates come from several sectors and interests in society—health, education, welfare, marketing, commercial, labor, and industrial. Officials responsible for planning and administering programs in many fields require more current detailed data, and legislators and administrators demand data to be used directly in allocating money and resources. Such demands for frequent small-area estimates heighten the impatience with data from the decennial census. Several methods, including the use of administrative records and synthetic and regression methods, have been developed for improved current estimates for populations of various small areas. The demand for increased detail, frequency, and accuracy of these data is increasing. The panel will explore this area of data requirements and report on its deliberations in the final report.

The congressional request for this study asked the panel to recommend methods for future censuses that provide an accurate and cost-effective collection of information deemed necessary for state and local needs. The issue of undercount and the differential racial undercount is one major area of study for the panel's future work. Within the general context of the demand for more accurate data, the panel will consider declining response rates to the mail questionnaire in relation to alternative data collection procedures that can meet the demands for accurate small-area data, taking costs into account. There may be methods for improving or dealing with mail response rates or using statistical methods to adjust for the final population count. The panel's final report will have additional discussion of the census undercount and census costs and will make recommendations for cost-effective methods that might deal with the undercount.

#### CENTRAL ISSUES AND CONCERNS

The Congress asked the National Academy of Sciences to undertake the study now being carried out by this panel because of congressional concerns about accuracy and budgetary costs of the 1990 census. Results of the 1990 census raised a number of serious problems and issues, both within Congress and among various groups of American citizens, especially, but by no means solely, those concerned by the large relative undercount of minority groups. At the same time, other groups—producers and users of federal statistical data and state and local governments, together with private firms heavily dependent on small-area census data—are concerned that whatever changes are made in census design or processes not only protect but improve the quality, breadth, and timeliness of data now available.

We have identified four major topics of concern: rapidly escalating costs, increased differential undercount of the population, increased requirements for accurate small-area data, and timeliness.

### *Rapidly Escalating Costs*

The cost of census activities has increased sharply since 1960. In 1990 dollars, the 1960 census cost about \$500 million;<sup>1</sup> the 1990 census cost \$2.6 billion, an increase of about 400 percent after adjusting for inflation (U.S. General Accounting Office, 1992). Population, and more importantly the number of housing units, increased over this 30 years, but even after making allowances for that, cost escalation has been severe.

The average cost of the census was less than \$10 dollars per housing unit in 1960 (in 1990 constant dollars) and was still only \$11 per housing unit for the 1970 census.<sup>2</sup> It escalated to \$20 per housing unit in 1980 and \$25 in 1990, an increase of 150 percent in real terms over 30 years.

Overall mail response rates were 78 percent in 1970, 75 percent in 1980, and 65 percent in 1990 (Bureau of the Census, 1990). If mail response rates dropped at the rate of change of 1970 to 1990, the U.S. General Accounting Office (1992:41) estimates the mail response rate in 2000 would be 55 to 59 percent. If the 2000 census is conducted using the same methods as the 1990 census, it would result in costs of \$4.8 billion in 2000 dollars (U.S. General Accounting Office, 1992:41).

### *Increased Differential Undercount of the Population*

If increased census costs had resulted in census improvements, particularly in better coverage and a decreased differential net undercount of minorities, it would be possible to argue the merits of costs versus coverage. However, both the overall percentage undercount and the differential undercount—the difference of the undercount between minorities and white populations—apparently worsened in the 1990 census (U.S. General Accounting Office, 1992:21–22). The net undercount of the total population, estimated from demographic analysis, rose from 1.2 percent in 1980 to 1.8 percent in 1990.

The difference in percentage net undercount between minorities and the white population increased in 1990. The data cited in the General Accounting Office report show the differential undercount between blacks and nonblacks. Our understanding, however, is that the proportions undercounted for the Hispanic, Asian and Pacific Islander, and Native American populations were also higher than for the white population in the 1990 census. The percentage difference in net undercount between blacks and nonblacks increased to 4.4 percent in 1990, the highest value since the Bureau of the Census began estimates of this type in 1940. The fact that spending more money did not produce a more accurate census was, in large part, at the center of criticisms leveled at the 1990 decennial census.

### *Increased Requirements for Accurate Small-Area Data*

The “one-person, one-vote” rulings of the Supreme Court in the 1960s and the Voting Rights Act of 1965 as extended and amended have substantially expanded the requirement for accurate population data, cross-classified by age and ethnicity, at the small-area level, for legislative redistricting and related purposes. And although the statutes do not specify the geographic level of detail that is required, census data provide it at the level of individual census blocks. Simultaneously over the past three decades, the number of federal statutes calling for the use of demographic and related data to apportion federal funds among states and localities has mushroomed. The statutes often do not specify the use of data based on the decennial census, but in practice the use of census data for these purposes is ubiquitous. There has also been a virtual explosion among state and local governments and private business firms in the development of computer data banks and computer models based on the use of census data at the block level for purposes of planning and operations. These uses have generated another set of demands for accurate data at small-area levels of detail. And, as we explore below, they have also raised important questions about relationships among federal, state, and local governments and private business firms with respect to the appropriate development and use of large-

<sup>1</sup> Total census costs are shown in U.S. General Accounting Office (1992:Figure 2.6) and in original source documents from the Bureau of the Census. We note, however, that Figure 2.6 of the General Accounting Office report is mislabelled as showing costs in constant 1990 dollars; the graph actually shows costs in current dollars. Total expenditures for the 1970 census, for instance, were \$222 million (Bureau of the Census, 1976b), which are shown in that amount in the General Accounting Office graph. In constant 1990 dollars, the amount is about \$800 million for the 1970 census.

<sup>2</sup> Census costs are calculated per housing unit because the census covers all housing units, including vacant ones. About 10 percent of housing units are typically vacant at the time of the decennial census, so the number of households is approximately 90 percent of the total number of housing units.

scale geocoding and geographic data systems, such as the Bureau of the Census's TIGER (Topologically Integrated Geographic Encoding and Referencing). TIGER provides a digital (computer readable) geographic database for the location and referencing of mailing addresses for small geographic areas for the nation.

These developments, the ways that the Bureau of the Census has responded to them, and the way that they are perceived by the public have interacted with each other both to raise the costs of taking the census and to generate increased perceptions of inadequacy in the resultant census statistics. Thus, for example, the depressed mail response rate is typically lowest precisely in areas or among groups that are most likely to be undercounted, leading to substantially increased costs in an effort to minimize the differential undercount and produce accurate data. The results, nevertheless, still fall short of meeting public demands for fairness and evenhandedness in redistricting legislatures and apportioning public funds. In turn, the Bureau of the Census over recent decades has sought to deal with the combined pressures of falling response rates and increased demands for detailed, geographically fine-grained and accurate data principally through the use of highly labor-intensive enumerative techniques for follow-up and coverage improvement. The current census design as it has evolved, can be contrasted for example, with increased use of sampling techniques as a means for follow-up when census questionnaires are not returned or as a means for adjusting the direct enumerative count.

Perceptions—not always in congruence with the facts—also have played a role in creating problems and raising costs for the census. Thus, as detailed below, what is required for fair and accurate redistricting and other uses are block-level data, which, when aggregated into relevant groupings, are reasonably accurate. Offsetting but unbiased errors at the block level, which will in any event always exist, need not contaminate the accuracy of the larger aggregates. If block-level errors are unbiased, the aggregation provides small relative error. Nevertheless, it is quite possible that realistic perceptions of public pressures for unachievable “perfection” in the count at small levels of geographic detail may have generated significant cost increases.

### *Timeliness of Census Data*

However accurate the initial data from the decennial census, they begin to lose accuracy the day after they are collected, especially at very small levels of geographic detail, as people change their residences and alter other characteristics collected as data by the census. This problem is more acute in a highly mobile society like the United States than in many other, more static societies. This mobility raises the issue of a potential tradeoff between the relative benefits of pushing for the highest possible degree of accuracy every 10 years versus some reduction of expenditures for the decennial census with the freed-up money used to improve intercensal demographic measures. These kinds of considerations have been part of the underlying rationale for some of the suggestions for alternative census designs—for example, a “rolling census,” which covers a fraction of the population each year, or a truncated census with a few questions decennially, supplemented by large intercensal surveys. The panel is considering these alternatives, along with other suggestions for different census designs; we summarize a few preliminary views below.

### PRELIMINARY ASSESSMENT

Our deliberations to date have led us to four general observations about the problems and issues outlined above. First, there are no conceivable changes in the collection of census data that will simultaneously meet all of the following objectives: overcome the consequences of a declining mail response rate through a census that relies principally on enumerative and labor-intensive follow-up techniques, provide detailed and reliable block-level data for redistricting and the Voting Rights Act, provide the other housing and demographic data widely demanded for cross-tabulation at the level of small geographic areas, and keep costs from growing rapidly. There is, in short, no magic bullet to current problems of the U.S. census.

Second, congressional concern with census cost escalation and differential undercount has led to a search for remedies for the 2000 census. One suggestion has been to reduce census content, perhaps by eliminating the long-form census questionnaire.<sup>3</sup> The panel intends to pursue the issue of census costs and burden. But, with

<sup>3</sup> Every person in the United States was asked during the 1990 census for some basic demographic information (race, age, relationship to household head, sex, and housing items for instance) on a short-form questionnaire. In addition, a one-in-six sample of persons was given a



one caveat, we do not believe that the content of the census is driving cost increases and coverage problems.<sup>4</sup> The panel notes that there are other reasons, expressed by congressional representatives, to reduce census content. The census has been perceived by some as unwieldy, and therefore subject to inaccuracies.

Some items on the long-form census questionnaire, such as those pertaining to income, may be subject to larger-than-average response error and require more follow-up verification. But the long-form questionnaire has remained at relatively constant length for the past four censuses, while costs have increased dramatically. Furthermore, the sampling rate for the long-form questionnaire has declined (from one in four in 1960 to one in six in 1990), which should have yielded lower overall costs. The panel appreciates the congressional concern about substantial cost escalations in the census; however, the panel does not find evidence that census content is the primary factor producing cost increases.

Although we believe it most unlikely that the addition or subtraction of a limited number of questions would significantly affect overall census costs, very large differences in the complexity and length of the census questionnaire might affect public willingness to cooperate, influence the response rate, and through that route affect costs. There is not much evidence on this point, but results from recent Bureau of the Census experiments with a very truncated form provide some insights. We will continue to pursue this point, although we again note the historical evidence, which strongly suggests that matters of content have not been a major driving force behind the recent and projected rises in census costs.

Third, from an inspection of recent trends in census costs, the panel has tentatively concluded that efforts to increase coverage, especially through highly labor-intensive enumeration techniques, are a key factor driving up costs. Moreover, efforts to improve coverage (i.e., the percentage of total population counted) have had relatively little absolute impact: coverage was better in 1980 than in 1970, but there was no such gain in 1990. Coverage improvement efforts have been carried to the point at which additional effort and expense may yield little or no improvement in either overall coverage or in decreasing racial differences in the net undercount. The panel is interested in exploring with the Bureau of the Census the components of cost increases from 1970 to 1990 and will report on this topic in its final report. Expensive efforts to improve census coverage are understandable given such forces as the impetus of the Voting Rights Act to provide detailed data on race and ethnicity at the block level. Nonetheless, it is appropriate to ask if this continued costly effort to improve coverage, so far unsuccessful, is necessary for future censuses.

Finally, as the discussion above makes clear, there are some major national policy choices involved in considering requirements and techniques for the decennial census. The panel sees its role as identifying those choices and providing analyses to make informed choices possible. Several examples illustrate those choices in what the 2000 and future censuses should look like. First, should the census be a highly intensive enumeration with a complete follow-up to all nonresponses or a reduced enumeration effort with more reliance on various statistical techniques for adjustment? The tradeoffs, in this case, involve costs, accuracy, and the acceptability of statistical adjustment. A second choice involves how to implement the basic requirements of court decisions mandating equality in the population sizes of voting districts. Both the court decisions and Voting Rights Act census data requirements are grounded firmly in the Constitution, including the Fourteenth and Fifteenth Amendments. Third, what level of information beyond basic demographic detail must be provided every 10 years for small areas? A large number of programs and funding allocation are based on decennial census data, but data from surveys or administrative records could be used to provide more timely estimates, albeit with less geographic detail than the census. To what extent can alternative sources meet reasonable needs, and at what cost?

In beginning our deliberations we found it necessary to examine the constitutional and other legal underpinnings for the various demands now placed on the census

---

more detailed long-form questionnaire with questions about income, schooling, occupation, and related social and economic items. Residents in areas with populations of 2,500 or less were sampled at a rate of one in two in order to obtain sufficient responses for accurate estimates of smaller areas. Persons in areas with 2,000 or more housing units were sampled at a one-in-eight rate. Special rules were made for persons in group quarters and those residing in areas where special enumeration was required. Overall, the sample design produced a sample of one in every six housing units in the nation.

<sup>4</sup> If a census were to be based on administrative records, as some people have suggested, then a strong case can be made that costs would increase to satisfy content demands. All current census data do not exist on administrative records and it would cost additional money to collect that information.



for demographic and other data. We first present a summary of our review of these matters to date; we then turn to a report—necessarily preliminary and tentative—of our consideration of alternative census designs and procedures.

### *Legal Requirements*

The panel has considered the constitutional requirements for the census. It appears clear that an attempt at a basic enumeration is required to meet constitutional requirements for reapportionment. Such an effort, in effect, is also required because of the need for small-area data for redistricting—including one-person, one-vote needs and Voting Rights Act requirements (see Appendix A for a legal and political history). Hence, a sample census (a census based completely on a sample) or a rolling census (with data collection over the decade) that does not include a complete enumeration once a decade are not reasonable options on legal grounds. However, we also considered these designs from a practical perspective.

The panel notes that there are practical distinctions and issues involved with proposals for a sample census or a rolling census. In a sample census, only some of the population is counted. In a rolling census design, different parts of the population are surveyed every year. Several variants of a rolling census design have been proposed.

One design proposes conducting a full census (including short-form and long-form content) of one-tenth of the nation's counties every year (Horvitz, 1986); a different set of counties would be covered every year of the decade. Horvitz claims that accurate data on internal migration could be developed from each year's one-tenth census for use in developing population estimates for the counties not covered that year. However, because the design does not provide for even a minimal census of the entire country at a contemporaneous point in time, it would fail to meet the constitutional requirement that reapportionment counts be based on an attempt at a complete enumeration of the entire population.

Another design variant proposes to survey a different one-tenth of the population each year, cumulating the estimates over 1, 2, or more years to increase their reliability for small geographic areas (Kish, 1981, 1990). Each year's sample would be nationally representative, instead of concentrated in selected areas. A one-tenth sample would provide state estimates of reasonable reliability, but we believe that use of such estimates would encounter the constitutional barrier to a sample census. Averaging the estimates from each year's sample over a 10-year period might be thought to increase their reliability, but these estimates would not pertain to a year chosen for reapportionment, but rather to the average experience of states and other areas over the prior 10 years. Because they are not based on an attempt at complete enumeration at a given time, they would, at a minimum, raise the question of whether they meet the constitutional requirement. Moreover, the "cumulated" estimates would be far more out-of-date with regard to the distribution of the population than would be estimates from a current full census.<sup>5</sup>

Designs that attempt to spread over a decade the collection of some of the information that is now obtained in the census merit evaluation as to their benefits and costs. They are worth considering if the benefits of more timely data and higher coverage would offset potentially higher costs. There appears to be a legal consensus, however, that such designs need to include a minimal complete census every tenth year to satisfy the constitutional requirement for reapportionment.<sup>6</sup>

### *Feasibility of a Sample Census*

Appendix A presents the argument that, for purposes of reapportionment, there needs to be an attempt to account for every inhabitant in the country. Accepting this argument, the panel concludes, therefore, that a sample census, no matter how large, cannot satisfy the constitutional requirement.

From the methodological perspective, it is not even clear how one would actually conduct a sample census. To obtain a sufficient degree of accuracy for reapportionment (and other purposes) would require a large sample,<sup>7</sup> that, in turn, would re-

<sup>5</sup> For reapportionment and redistricting in 2001, 2011, and so on, census data would be about 1 year old, while the cumulated estimates would be centered mid-decade, making them 6 years old. See Fellegi (1981) for a detailed critique of the Kish proposal, including conceptual and operational aspects.

<sup>6</sup> The Bureau of the Census is currently evaluating "continuous measurement" designs that include a year-zero census together with "rolling surveys" throughout the decade to obtain long-form items and update short-form items (see Alexander, 1993). See discussion below for additional material on continuous measurement designs.

<sup>7</sup> With no disproportionate sampling by size, a sample as big as the 1960 long-form sample—about one-fourth of the population—would be required to provide estimates for the smallest

quire the development of a frame from which to draw the sample. In the census, the frame would be the list of addresses that is used to mail out census questionnaires.<sup>8</sup> In addition to developing the address list, it would be necessary to carry out activities to determine the list's accuracy for purposes of obtaining a proper random sample of people, as the list would likely include nonresidential addresses (e.g., businesses) as well as exclude some residences or contain a systematic bias against larger or smaller households. One such evaluation activity might be to conduct a second independent survey either before or after the sample census itself, which, at a minimum, would have to ascertain the number, age, and sex of respondents. As a result of all these activities, the Bureau of the Census would quickly wind up, for all intents and purposes, conducting virtually a complete census.

A sample census, assuming that it could be conducted, would face other problems. All the available evidence shows that household surveys experience higher net undercoverage rates and more severe differential undercoverage than does the complete-count census.<sup>9</sup> The coverage errors in a sample census could be evaluated, but any adjustment process would make a larger contribution to the final estimates than would be the case for a complete census. Another problem that could contribute to coverage errors is the likely difficulty of publicizing a sample census in which there is no intent to try to contact everyone. Finally, there would appear to be little likelihood of achieving significant cost savings over a complete census since the full cost of creating the address list would be incurred. Moreover, the need for reliable small-area estimates would preclude the clustering of field operations as is typically done in smaller household surveys.

#### CONSIDERATIONS FOR EVALUATING ALTERNATIVE CENSUS DESIGNS

This section is divided into two parts: some general issues common to many alternative census designs, including the current one, and some preliminary considerations about specific possible new census designs.

##### GENERAL ISSUES

##### *Sampling for Nonresponse Follow-Up*

The constitutional requirement for complete enumeration does not necessarily rule out the increased use of sampling as part of a census. Sampling has several potential uses that are worth inspection for increased use in future censuses, and there needs to be a distinction between the various uses of sampling in a census. One use is to sample the population and to estimate the total population: in such a sample census, not every person is enumerated. However, there are two broad other uses of sampling for a complete count census: sampling for nonresponse and sampling to estimate undercoverage or errors.

Sampling for nonresponse involves follow-up visits with a selected proportion of nonresponding addresses and then inferring the information that would have been collected for all addresses which did not return their mail questionnaire. The Bureau of the Census has made use of sample surveys to estimate undercoverage and errors for several censuses, although the survey sample size would need to be larger if survey-based estimates of undercoverage were to be incorporated into the census. The panel intends to examine the use of sampling to adjust for undercoverage and differential undercoverage. Such use of sampling involves a crucial tradeoff

---

states that (say, with 95-degree confidence) varied by no more than 0.5 percent from the estimates obtained from a complete census (calculated from the formula provided in Bureau of the Census (1983a:27), assuming a total U.S. population of 250 million and a population of 0.5 million for the smallest states). If the constraints on precision for the smallest states were relaxed to 1 percent, the sample size could be smaller (about 1 in 13 if there was no disproportionate sampling). The size could be smaller yet, if the sample was designed specifically to provide about the same level of precision for all states; however, smaller samples would not likely serve the interrelated purpose of redistricting, which requires data for very small areas.

<sup>8</sup> This line of argument does not necessarily apply to all sample surveys, many of which can make use of area frames. However, a list frame is necessary for a large-scale survey that is intended to provide estimates of acceptable reliability for all small areas.

<sup>9</sup> For example, in March 1986, the Survey of Income and Program Participation and the March Current Population Survey covered only 80-2 percent of black men aged 16 and older and 93 percent of nonblack men aged 16 and older. Coverage ratios were somewhat higher for women. (Coverage ratios compare the estimates from a survey, using the initial survey weights that take into account the sample fraction and household nonresponse, with the corresponding census-based population figures not adjusted for undercount; see Citro and Kalton, 1993:Table 3-12; see also Shapiro and Kostanich, 1988.)

between expensive efforts to improve coverage everywhere and sophisticated designs using samples to provide data for adjusting the direct census counts.

Although a sample census appears precluded on constitutional and methodological grounds, the panel believes that the spirit of the constitutional, legislative, and judicial history regarding "enumeration" is compatible with the use of sampling as part of the census process, so long as that process includes an effort to reach all U.S. inhabitants. Specifically, the panel believes that census designs that use sampling for the follow-up stage of census operations (after an initial attempt has been made to give a questionnaire to everyone) and for coverage improvement programs (including adjustments based on sample surveys) would both meet the data requirements for reapportionment and have the potential for increasing census accuracy for this purpose while reducing census costs.

There are precedents in previous censuses for the use of sampling (see Appendix A). Several court cases have explicitly upheld the constitutionality of an adjustment based on a survey (such as the postenumeration survey in the 1990 census), citing the importance of having data as accurate as possible for reapportionment and redistricting.<sup>10</sup> The question of the legality of sampling for follow-up to nonresponses has never been explicitly raised in the courts; however, language used in the court cases just cited would clearly seem to be consistent with its use (see Appendix A for a history on this issue).

There are sizable census costs for following up all nonrespondents to the decennial census mail questionnaire. Moreover, many of the "nonrespondents" are vacant dwelling units for which a personal visit is needed in order to ascertain the vacancy status. For the follow-up procedures, two main types of savings might be considered. One approach would be to make greater use of the Postal Service for reporting on the vacancy status and not schedule a census visit unless there are reports that the unit is inhabited. A second approach would entail taking a sample of the nonrespondents and, using statistical techniques, estimating the population data for all nonrespondents.

The use of sampling for following up nonresponses could provide considerable cost savings. The Bureau of the Census estimates that sampling 50 percent of nonrespondents in the 1990 census could have saved an estimated \$215 million (Bureau of the Census, 1990). Even smaller sampling fractions could produce progressively greater savings: \$325 million with a 33 percent sample, \$435 million with 20 percent, and \$460 million with 10 percent (all figures are in 1990 dollars). If 10 percent of nonresponses were sampled, the Bureau of the Census' estimates suggest a possible saving of about 18 percent of total 1990 census costs. Sampling nonrespondents could provide one of the largest single sources of cost savings in census operations. We note, however, that these cost estimates are not based on operational experience, and there is not sufficient information for evaluating the usefulness of sampling for nonresponse follow-up.

More precise information is needed on the costs and benefits of sampling for nonresponse follow-up in the census. A number of issues require thorough investigation: for example, the extent of correlation between sampling and nonsampling errors, the cost savings that would be reasonable to expect from different levels of sampling, and the implications for other important census data requirements. The panel believes that it is important to research these issues on a timely basis. In order to obtain the needed information, it is important to examine this methodology in the 1995 census tests. Only such testing can provide the requisite data needed for serious consideration of this technique for the 2000 census. Without such testing, there would be no information for discussion of its possible inclusion in the 2000 census design.

**Recommendation 1.** The panel recommends that serious consideration be given to investigating sampling for nonresponse follow-up in the 1995 census tests. The panel recommends testing to provide information on the costs, effects on small-area data, and statistical problems.

#### *Accuracy of Block Data*

As noted above, some people have argued that the key limitation of data from a sample census would be that it would not provide a high enough level of accuracy for small-area data. But even with a complete count, there are inaccuracies in the coverage and the content items of the censuses. No data are perfect. The important

<sup>10</sup> The recent case of *City of New York v. U.S. Department of Commerce* (1993, Eastern District, New York) is the latest court decision that discusses, in part, the statistical methods that might be used for the conduct of the census: Appendix A presents a fuller discussion of legal issues.

concern for small-area data is not to assume that they are perfect, but to detect errors and to be able to take them into account, in analysis or in legal arguments. At present, however, information is lacking about the nature of errors in small-area data, including problems arising from both errors and biases. Especially lacking is the type of information that is routinely available for larger geographic areas: for example, data are available on the nonresponse and imputation rates, by census item, for states and major metropolitan areas. Estimates of errors are not necessarily needed for every single census question, nor obviously for every single small geographic area. However, overall error information is needed for small geographic areas (e.g., census blocks) in order to make a proper assessment of the accuracy of the data.

Over recent decades, there has been a movement toward using information about increasingly smaller geographic areas. About 50 years ago analysts began using census tracts, which have a population size of 4,000–5,000 people. Demands to the Bureau of the Census for data that could be used for various administratively defined areas have more recently resulted in the availability of data by census block. Blocks are the smallest geographic unit that the Bureau of the Census uses for release of census data; data for about 10 million census blocks were provided in the 1990 census. The panel is concerned about the lack of information about the accuracy of small-area data, with the observation that some users assume that the data are “perfectly accurate.”

Although census data users need block data as the fundamental geographic units for analysis and for constructing alternative geographic units, it is important to recognize that there cannot be total accuracy for block data. Indeed, census block data will never possess perfect accuracy. The key point for such uses as the Voting Rights Act is accuracy of the aggregation of census blocks to form districts. It would be useful to have more careful statistical analysis and error profiles on block-level data, and successive levels of aggregation, so that the public debate can be better informed. Analysts could use estimates of gross errors, including sampling and non-sampling errors, at the block level at the time of collection and other estimates of population change over the decade to inform the courts and other data users about the problems of focussing on absolute population equality in redistricting—to the exclusion of such other criteria as compactness and contiguity.

The Bureau of the Census routinely provides information about the accuracy of census data, from the content reinterview surveys and from the postenumeration surveys. The panel believes that it would be helpful for census data users to have information about the accuracy of small-area data. Prior to the 1990 census, error analysis at the block level could not be carried out; however, the design of the 1990 Post-Enumeration Survey makes such analysis possible.

Recommendation 2. The panel recommends that the Bureau of the Census analyze 1990 Post-Enumeration Survey data to produce estimate of gross errors at the block level in the census. This information is required for examining census requirements for the accuracy of small-area data in the future.

### *Timeliness*

The panel is concerned with the issue of timeliness of both short-form and long-form items. Even if the coverage and content items of the decennial census were perfectly accurate at the time of collection, major discrepancies would occur during the 10-year period of their use if they were used without updating. In a country which had about 10 million new immigrants enter during the 1980–1990 decade and in which about 20 percent of the population changes residence annually, any coverage defects of a census are very shortly dominated by the effects of population mobility. In addition, of course, changes in unemployment, poverty, and other socioeconomic shifts also occur after the taking of a census. Nevertheless, social and economic data from the census are used until the next census—up to 12 or 13 years in some cases—for many small-area and national population estimates.

Intercensal population estimates have been made by the Bureau of the Census for many years. For example, the intercensal population estimates program provides annual estimates for states by age and sex. The Bureau of the Census's estimation program relies on several administrative record systems, including birth, death, and immigration information, as well as Internal Revenue Service tax-filer data as a source of information about internal migration. Population estimates for states, counties, and cities are made by many state and local agencies. These population estimates often rely on nonfederal administrative records.

Cumulated or rolling census designs could help to provide more timely information than a decennial census, although the costs are likely to be higher than the current census design. Furthermore, as noted above, the panel does not believe that



a rolling census design would provide the constitutionally required complete enumeration of the population unless it included a minimal census once a decade.

Another possible way to improve timeliness is for the federal government to invest in a mapping and geographic coding system, such as TIGER, and work with states and localities to encourage them to incorporate their administrative records in this system. Federal records could also be linked to the system, but it would be important to have a partnership with states and localities in order to maintain the geographic database, the housing inventory, and administrative record information that is maintained only by state and local governments. Over time, administrative records linked to a geographic reference system could provide more frequent data and possibly substitute for some current census items (especially housing data). Appendix B outlines the development of a possible system for consideration by the federal government.

The panel would like to emphasize that a full account of "errors" in the use of census data involves two main sources. One source is actual census errors, which have received a predominance of the public criticism and legal wrangling. The other source is population shifts. It is the second source of change that produces, over the duration of a decade, the major cause of discrepancy between the original census information and the phenomenon that it is supposed to represent. With heavy immigration and substantial migration, population shifts dwarf any errors that existed in the original census information.

#### *Small-Area Data*

Small-area data are used for a variety of purposes, including the allocation of federal and state funds, public and private planning, determining the eligibility of a locality for funding or government programs, and scholarly research. However, except for population and income estimates for larger geographic areas, intercensal program and policy decisions rely on decennial census information.

The particular challenge of estimates for small geographic areas is that they can experience rapid population change: the smaller the geographic area, the greater the influence of migration and the possibility of rapid population changes. A small area can, for example, over a period of a few years, experience the expansion of a new suburban development. Or, over an equally short period of time, a group of immigrants might move to a small neighborhood of a city. In each case, decennial census information may no longer offer adequate information about the number or characteristics of the people in that small area.

Similar observations can be made for small population groups in the United States. Information for smaller ethnic groups is based principally on data from the decennial census. During periods of heavy and shifting immigration, such as has been occurring during the past 20 years, decennial census information presents an inadequate picture of the number and socioeconomic characteristics of small ethnic groups within 4 or 5 years after their arrival.

Some organizations, such as private for-profit companies, have increasingly developed alternative sources of information for making decisions that require small-area data. For decisions about site selection, advertising and promotional campaigns, and market research, companies have increasingly acquired and generated transactional databases that provide frequent small-area estimates. Transactional data include any event that is routinely recorded by businesses and can be linked to an individual or household, such as grocery store purchases or checking account monthly balances. Some transactional data are now generated from cash registers and inventory controls for real-time estimates of change.

Transactional data now provide frequent small-area estimates for many businesses. The panel has not examined work that reports on the quality and usefulness of these private estimates for small areas: the data may be frequent, but they may also be of poor quality. The panel is concerned, however, with the potential for improving the timeliness of small-area estimates, especially for their usefulness for the public sector.

#### *A Geographic Database*

The 1990 decennial census in the United States relied heavily on its geographic database, TIGER. All indications are that the Bureau of the Census plans to rely on a TIGER-type system for the 2000 census, regardless of specific census design. There are two separate files in the present form of the Bureau of the Census's geographic address system. One is TIGER, a cartographic database with physical features and address ranges. The other is a separate database of specific address lists, so that TIGER does not itself reveal individual housing addresses (nor any information about the occupants). TIGER can be linked to the address list for use in census planning and operations.

Although TIGER was used for the 1990 census, several questions arise about the features of a geographic database for the next census and beyond. Can TIGER serve as the basis for an enhanced system, or is a new system required? If a geographic database is to be used only for the conduct of the decennial census, what are the required features of the database? Is a geographic database, such as TIGER, critical for developing intercensal estimates? If so, then the geographic database needs to be maintained throughout the decade. These are important questions to raise about any geographic database, including TIGER. Also, the panel knows very little about the cost of TIGER or alternative geographic databases. The cost issue needs to be explored as well for any recommendations about the geographic underpinning of census activities.

The panel views geographic information as having several important objectives for census activities. First, there is potential use of administrative records for small-area estimates during the intercensal period. Some records have a geographic reference and can be geocoded<sup>11</sup> without reference to other data. But for small-area estimates (and for linkage to other records), a geographic referencing system such as TIGER is needed. This is an important activity (discussed more fully below), and the panel supports efforts to provide small-area estimates more frequently.

A complete address and geographic reference database appears to be critical for all designs currently under consideration for the U.S. census, although a modern census can be taken without a prepared national address file and linked geographic reference system. In one approach, that taken in Canada, enumerators deliver a census questionnaire to each household, noting the addresses to which the questionnaires were delivered. Essentially, the address list is created at the very time of the census. In the United States, however, an accurate address and housing inventory, located geographically, is required for decennial census operations. The inventory is also important for state and local cooperation with the census because agreement on the census count of housing units is a major source of controversy for the overall population count. This is a second objective and the panel supports this effort.

A good geographic database is also important for many users. The availability of the Bureau of the Census's TIGER database and its enhancement and support by private companies and federal and state agencies have opened the door for the widespread analysis of small-area data. Moreover, recent technological achievements, including more powerful microcomputers and large-scale data storage on CD-ROMS, have distributed small-area data to many new users. These users will be unwilling to retreat to an earlier, more limited type of census data. The panel supports efforts to improve small-area data and to have the data and its geographic referencing available to a wide variety of users.

In the panel's initial deliberation, it has realized that some form of address list linked to a geographic system is important for the census. In its future work, the panel intends to consider numerous alternative ways of handling geographic referencing for the census, with consideration of cost, efficiency, and equity.

#### SOME POSSIBLE NEW DESIGNS

There are a variety of ways in which a population can be counted. A conventional modern population census attempts to provide a count of the people within a territory at one point in time. Although there are many variants of specific procedures, the modern population census as used in most countries today involves seven steps:

- (1) a definition of the population to be considered as being in scope for the census,
- (2) determination of the content to be included on the questionnaire based on an extensive examination of users' needs,
- (3) careful testing of alternative questionnaire wordings and formats,
- (4) a systematic preparation of lists of dwellings in which the population lives,
- (5) the hiring and training of a group of enumerators,
- (6) the use of enumerators to question the inhabitants, either all of them or only those who did not satisfactorily complete their questionnaires, and
- (7) the processing and analysis of census questionnaires.

For the counting of populations, most countries depend on enumerators to either visit households or to compile results from mail questionnaires. As noted above, some countries, including Canada, have enumerators deliver questionnaires directly

<sup>11</sup> Geocoding means providing a geographic reference for administrative records. To do this, street addresses or address ranges are needed for the records. In the absence of addresses or address ranges, a procedure for assigning the records to geographically defined areas is necessary.

to each household. The U.S. census operation relies primarily on an accurate mailing address for every dwelling unit and then uses the U.S. Postal Service for delivery of a mail questionnaire. If questionnaires are not returned, the census office directly contacts the nonrespondent household. The relative merits of having enumerators delivering the census questionnaire versus exclusive reliance on a mail questionnaire, which requires an accurate mailing list prior to the census, is a topic that the panel will continue to study.

Even for the modern population census, there are variants of specific procedures and widely different levels of intensity for achieving coverage. Some countries distribute their census questionnaires, do minimal follow-up, and accept some net undercount in doing a less expensive census. In contrast, the U.S. census operation is a very intensive attempt to contact every dwelling unit and obtain demographic information for every person in the household. As part of this approach, the U.S. census operation involves an intensive effort to prepare an accurate address list, is active in the field for a longer period than census operations in other countries, and attempts to follow up all nonrespondents. As a result, the U.S. census is now very expensive. One major census design alternative, therefore, is for a less intensive, cheaper census. Such a census need not necessarily be a poorer quality census, however. The use of such statistical techniques as integrated coverage improvement or postenumeration survey census adjustment could maintain the overall quality of census coverage.

#### *An Administrative Records-Based Census*

There is another main class of census designs that might be considered for use in the United States: the use of administrative records, possibly linked to a population register. The panel notes that some countries use this approach, but has not considered all the merits and limitations of such an approach for the United States. We note in this interim report some features of a register system, but without making any recommendations for the use of registers in the United States.

Some countries have inaugurated a continuous system of "population accounting" for census activities. In such a system of continuous registration, the location of every person in the country is known at all times. Its major characteristic is that some sort of personal record or identification is required. The advantages of continuous registration—from the demographic perspective—is that it provides full information about the population whenever it is required. And, when working well, it eliminates the need for a census by enumeration. The disadvantages are perhaps less immediately obvious but are nevertheless important: the system is laborious to operate, costly, liable to accumulate errors, and lastly, but by no means unimportant, may be seen by the public as an inappropriate invasion of privacy.

In spite of the disadvantages, continuous population registers have been successfully developed in several nations, including Belgium, the Netherlands, Switzerland, and the Scandinavian countries. The Swedish system has its roots in the seventeenth century, and the Dutch register has been operating for more than a century. A similar system was brought into force in the United Kingdom during World War II but was discontinued thereafter.<sup>12</sup>

During the past decade, several European countries have linked additional administrative records to their central population registers and have discontinued direct inquiry censuses. Denmark, Finland, and the Netherlands now rely on population counts from their central records, although they supplement their administrative records with sample surveys. Other countries, including Sweden, now use administrative records to provide information that, in turn, does not need to be asked in the census. A critical question is whether it is desirable for the United States to consider developing methods for a population census count that use administrative records, but without the support of a continuous population register.

To consider this broad option in brief terms, the panel believes that administrative records cannot be adapted in the near future (for the year 2000) to meet even relaxed demands for detailed demographic information for small geographic areas. No federal or state administrative systems now possess race and ethnic data for the U.S. population, combined with coverage of all households and accurate information on current address. As a substitute for census data, it would be difficult to redesign current administrative data to meet the Voting Rights Act requirements for data on race and ethnicity for small geographic areas. Another limitation of existing administrative records is the paucity of demographic and socioeconomic variables. The panel agrees with a letter report from the Panel to Evaluate Alternative Census

<sup>12</sup> See Redfern (1989) for a review of the use of administrative records for a census, including a summary of activities in Europe.



Methods to the director of the Bureau of the Census, dated December 14, 1992, which finds that it is not feasible to consider the use of an administrative records-based census for the year 2000.

It is important to look beyond the 2000 census for an evaluation of a possibly more effective use of administrative records. Recent research suggests that a high proportion of the U.S. population can be counted through existing administrative records (Sailer et al., 1992). The coverage of administrative records is likely to expand in the future. Moreover, if it became important to use records for census purposes, efforts could be made to expand the content and improve the quality of the records. An administrative records census may be a real alternative in 2010 or 2020, but it will not be an alternative if work does not begin soon in order to accumulate experience in working with the records. This is an important area for further consideration, and more research is needed.

#### *Key Issues with Administrative Records*

Consideration of the use of administrative records as a complement or substitute for the current census must address several issues. Although the coverage of some administrative record systems is high and expanding, no single record system covers the entire population. Two examples provide an illustration. The Internal Revenue Service collects information from its annual tax returns and related documents; however, it is a challenge to produce estimates for nonfilers, who are not in its administrative records. As another example, the Social Security Administration maintains a register of social security accounts. Increasingly, infants are registered with the Social Security Administration at birth and, since immigrants need to obtain a social security account in order to work, the country is moving toward a universal registration system. However, there persists significant illegal immigration to the United States, with about 200,000 illegal entrants annually and approximately 2 million current illegal residents (Bean et al., 1990). Illegal immigrants typically use fraudulent social security documents or other people's account numbers. There is a need therefore to assess differential coverage for administrative records systems and to examine whether an adjustment procedure could be developed to use administrative records for a census count.

Currently, race and ethnic identification information needs to be provided by the census. There are no race and ethnic data in Internal Revenue Service records, and Social Security records do not contain such information for everyone. Birth registration certificates include the race and Hispanic status of the mother (and sometimes of the father), but the racial information is kept confidential by states' vital registration systems. Moreover, there are questions about the racial and ethnic identity for the many children who are born to parents of different or mixed race or ethnicity. Can one presume the race or ethnic identity of an adult on the basis of information from a birth certificate? The United States has continually changed its sense of race and ethnic classification over time. Even if a record system included racial and ethnic information for the entire population, the categories may change (as was done during recent censuses when a category for Asian and Pacific Islander was added), and the entire population would need to be resurveyed. Finally, there are serious concerns about the different reporting of race and ethnicity in a nonthreatening census context and in administrative records.

Beside the challenge of obtaining essential race and ethnic information from administrative records, there are serious doubts that current records (from either the federal, state, or local governments) could provide most of the information currently collected in the long-form census questionnaire. Could records be redesigned to collect this information? Would large surveys have to be used in conjunction with an administrative records-based census?

A final issue for examination is the accuracy of addresses for individuals in administrative records. The census needs to allocate individuals to fairly small areas. This can be done if individuals can be assigned to a specific dwelling unit, then placed within census blocks. However, this assignment requires that administrative records are current and have an accurate street address. In some cases, administrative records that lack up-to-date, accurate addresses could be linked to another record system that does provide such addresses. The challenge would be to link several individual records at the household level. These are issues for the panel to examine in its future work.

#### *Further work on Administrative Records*

The panel concurs with the recommendation of the Panel to Evaluate Alternative Census Methods (1992:5-6):

Recommendation 3. The panel recommends that the Bureau of the Census should (a) initiate a separate program of research on uses of administrative



records, not directly related to the 2000 census, focusing primarily on the 2010 census and on current estimates programs; (b) undertake a planning study to develop detailed design options for a 2010 administrative records census; (c) seek the cooperation of federal agencies that maintain key administrative record systems in undertaking experimental minicensuses and related projects based on administrative records; and (d) give priority to some use of administrative records in the 2000 census for those purposes for which such usage is feasible, such as coverage and content improvement and coverage evaluation.

Title 13 of the U.S. Code governs the Bureau of the Census's mandate for data collection and how it distributes statistical information. The requirements of Title 13 limit the ability of the Bureau of the Census to share information. Expanded work with administrative records needs to take into account problems of data access for use by the Bureau of the Census and availability of data for linkage to other records. Expanded use of administrative records requires interagency cooperation. Some records may require redesign. Some records may need improvements for accuracy, including information on residential addresses for proper allocation to the dwelling unit. And access to other records may be required. All of these issues pertain to the participation of the Bureau of the Census and other agencies in teamwork to expand and improve the use of administrative records for census use.

There is also another Title 13 concern for the census. The current provisions in Title 13 produce a "one-way street" for data exchange with the Bureau of the Census. For example, Title 13 requires the Bureau of the Census to safeguard the addresses of housing units. Although the Bureau of the Census originally developed its mailing list from public lists and in association with the U.S. Postal Service, it maintains the confidentiality of the address list itself. One reason the Bureau of the Census does not release the list of addresses it finds is that the list may contain illegal addresses from the perspective of local government housing law. There has been fear that the local government might use a list to enforce housing regulations. This one-way street has become a serious problem for decennial census operations. City and local authorities often provide the Bureau of the Census with their lists of housing units; the Bureau of the Census then comments on overall discrepancies, but does not allow local authorities to scrutinize the census list. This fosters suspicions on the part of local authorities. It also does not allow the Bureau of the Census to work in full partnership with local officials for the improvement and reconciliation of address lists.

As an alternative, one might imagine a national housing register—a listing of housing and addresses only, with no individual or family information—that is maintained in collaboration by the Bureau of the Census and local officials. Such a list would be referenced to its correct geographic location and would be available for intercensal use and would, by its very nature, provide a continuous inventory of housing by small geographic areas. It would also be public (or available to local officials with restrictions for its use) and would avert one of the major disagreements of local officials with the decennial census: debates about the correct number of housing units by small geographic area. Such a reconciled, geographically referenced housing list would also improve the quality of the decennial census count.

In sum, the panel believes that an organized program is needed to explore the potential use of administrative records for several purposes. One purpose is as a substitute for the decennial census itself. From the experience of several European countries, it is clear that administrative records can be used in conjunction with a population register to provide a substitute for an enumerator-based census. It is less certain whether administrative records, relying on an accurate geographic-housing database, could be used by themselves for a population census. A second purpose is to explore the ways in which administrative records could complement a census, for coverage improvement and for a contribution to content. Finally, the panel believes it is useful to examine the possible use of administrative records for intercensal estimates (see below). Administrative records are used already for intercensal estimates. To give one example, the Bureau of the Census relies on changes in annual taxfilers' addresses to provide internal migration information for intercensal estimates. There is a natural link of intercensal estimates to the census: intercensal estimates can provide more timely data than the census and, therefore, decrease reliance on the census for small-area data. Moreover, experience with administrative records for intercensal estimates provides valuable knowledge about the limitations and merits of records and improves the records themselves.

#### *Administrative Records for Intercensal Estimates*

As mentioned above, redesigned administrative records have the potential to complement current census data collection procedures. Also, there is a possibility for a

census that would be primarily taken through the use of administrative records. We recommend (see above) that an important first step in examining administrative records is to begin working with them now. If administrative records are to have an expanded use in the decennial census, then there is an urgent need to start to exploit them more heavily for intercensal estimates. Experience is needed for not just tabulating available administrative records, but to gain experience in imputation, adjustment, and interagency coordination. Much of this experience could be gained by starting work on administrative records for the provision of intercensal estimates. Accelerated use of administrative records for intercensal estimates would encourage the Bureau of the Census to work out arrangements for obtaining important records that are outside the Bureau of the Census. And, although the Bureau of the Census has been using administrative records for years, their expanded use for intercensal estimates would provide the necessary experience that is needed for assessing their potential for the decennial census.

Another rationale exists for using administrative records for intercensal estimates. Census data are available only every 10 years. It usually takes 3 or more years before small-area data estimates are available from the census. This time lag means that much of the nation's small-area data are 3 years old when first available, and they are used until they are about 13 years old. On average, then, U.S. small-area estimates are approximately 8 years old over the decade of their use. Administrative records have the potential to provide much more frequent information for small geographic areas, on important variables such as population and housing counts, poverty, and income.

### *Continuous Measurement*

Continuous measurement designs for the census propose to collect information throughout the decade. In one design, a rolling census, the entire data collection occurs over a 10-year period. In another design, a periodic minimal census could be combined with various forms of continuous measurement, such as in the case of the design suggested by Herriot et al. (1989). This design would include a complete census with minimal content every 10 years, followed by a survey every year of one-tenth of the states, with a sampling rate similar to the census long form—one-fifth or one-sixth of the population in the sampled states. The merits of such a design require careful scrutiny, but it would clearly satisfy the constitutional requirement.

There are various proposals for a census based on data collection throughout the decade (Herriot et al., 1989; Horvitz, 1986; Kish, 1990). The common element of these designs is that census data would be collected continuously in time and space, with only a minimal decennial census (if any). These designs offer an amalgam of approaches, usually with unstated objectives. As argued above, the panel does not believe that a rolling census design would fulfill minimal census requirements unless it includes a complete enumeration at one time. Nevertheless, rolling census designs include some features for more timely data that are worth reviewing.

In one possible collection scheme (adapted from Herriot et al., 1989), the census would be a decade-long program with four components. The first component would be a minimal census with a simplified short-form questionnaire. This component would provide information for reapportionment, redistricting, the Voting Rights Act, and for the decade's sampling scheme. The second component would provide population and housing characteristics for small geographic areas through a sample of one in five households over the decade, with states rolling in and out of the sample on an annual basis. The third component would be a large ongoing national survey that would provide estimates for geographic areas with 100,000 or more persons. This survey would offer timely estimates for all states, major metropolitan areas, and large cities on a variety of demographic and housing topics. The fourth component would be an expanded intercensal estimates program that would provide monthly population estimates for all states, all counties, and all cities of 100,000 or more population.

The key feature of continuous measurement designs is that the population would be surveyed throughout the decade, rather than at a single time. Proponents of continuous measurement designs argue that the quality of data could be better than in the decennial census because data collection would be done by a permanent staff, who would be experienced at collecting census data. The census workload would be distributed uniformly over time and space, and a trained, permanent staff would be responsible for all data collection.

Continuous measurement approaches have some noteworthy advantages. Although a uniform core questionnaire would be used for all survey operations, state-specific questions could be included. Also, the large ongoing national household survey could give monthly and quarterly estimates for congressional districts and

most states. And quarterly estimates for all large metropolitan areas could be available from such a survey.

Costs for continuous measurement data collection would be spread out over a decade, which would reduce the marked census costs for a limited period of time (although reductions in annual appropriations might jeopardize the overall enterprise). Precise cost estimates are not available for the various proposed continuous measurement designs, although it does appear that this type of design would cost much more than the 1990 census (Herriot et al., 1988).

#### KEY ISSUES IN CENSUS REQUIREMENTS

This section briefly presents some key issues that the panel will consider further in its work.

##### *Racial and Ethnic Data*

An accurate count of the population by race and ethnicity, including individuals' locations for small geographic areas, is critical for the requirements of the Voting Rights Act. The U.S. Office of Management and Budget (OMB) sets minimal categories for race and ethnic questions, although the decennial census has collected more detailed information than is required by OMB Directive 15. Decennial census data are routinely used for analysis of equity issues for minority groups. Thus, an additional requirement for race and ethnic data is that they can be cross-classified by other social and economic information.

Two issues for race and ethnic data are important for the panel's further study: coverage and classification. Some race and ethnic groups have very small numbers, relative to the total population, and they are understandably concerned about possible undercounting. The classification of race and ethnic groups has evolved over many censuses. In addition to the race and ethnic information required by Office of Management and Budget Directive 15, the 1990 census also collected more detailed information on categories within the Hispanic, American Indian and Alaskan Native, and Asian and Pacific Islander groups. However, there may be substantial costs to using the decennial census to obtain finer distinctions—costs for data collection as well as the costs to accuracy of too many categories.

In recent censuses in the United States, the person in each household who completes the census form identifies the race and ethnicity of all persons in the household. By definition, whatever response is recorded is an "accurate" response. Although it might be the case that if another person in the household were the respondent, the racial and ethnic classifications of the persons in the household would be different, it is nevertheless the case that the respondent's classification of his or her race and ethnicity and those of other household members is regarded as accurate irrespective of objective reality (however defined).

It is clear that the actual wording of race and ethnic questions can affect responses: for example, having an explicit box to check for Asian self-identification results in more persons who classify themselves as Asian than does an open-ended question to which respondents can write in the classification "Asian". Context also matters, so that the same person might respond differently to the same wording of questions when asked in the census and when asked, for example, in a job application form.<sup>13</sup> Different classifications might be expected when one compares self-identification in the census with information on, for example, a birth or death certificate when the wording of the questions is not the same and when the information on the certificate might be provided by a third-party observer (such as a nurse or funeral director). Such different classifications make it difficult to assess the quality of a census that uses demographic techniques, in which counts in the census are compared with expected counts based on past birth and death records and estimates of international migration. They also pose problems for a census based on administrative records, since conflicts among racial and ethnic classifications in different record systems would have to be resolved and since there is little guarantee that the result would correspond to the classifications that would be obtained from the census as currently conducted.

<sup>13</sup> Fellegi (private communication) provides an example of this very great sensitivity to race. Racial self-reports were obtained from the Canadian Public Service Commission, the federal government's central personnel management agency, and compared, in one case, to the 1991 Census of Canada and, in the other case, to an employee survey. With results directly compared, there is evidence that individuals may make rather different reports of their racial status in the nonthreatening, anonymous context of the census. In the Canadian data, individuals were more likely to report minority race status in the census than in the other two surveys.

This problem would appear to pose little difficulty for any census design methodology other than one based on administrative records. Nonresponse follow-up of a sample and imputation of characteristics of those not in the follow-up sample would be based on answers to the same questions contained in the census. Likewise, a post-enumeration survey that allows for statistical adjustment for those who are neither counted in the census nor in the nonresponse follow-up sample and not imputed on the basis of the nonresponse follow-up sample would use the same questions on race and ethnicity contained in the census.

The real problem concerning race and ethnicity is that minority groups—especially blacks and Hispanics—are less likely to be counted in the census than are whites and are less likely to be found in any attempt to enhance coverage. The challenge for a statistical adjustment, therefore, remains the same regardless of other features of the census design: to estimate accurately the undercount of such minority groups relative to the white majority population and to place those persons in the correct geographic location. This challenge is explicit for redistricting—for which the counts of different racial and ethnic groups are important for the actual drawing of the boundaries for House seats. The challenge is also implicit for reapportionment—for which the race or ethnicity of those included in the census counts does not matter per se so long as states are enumerated to the same degree, but this goal would be unattainable so long as the racial distribution is not the same across states and undercounts differ by race.

The panel plans to examine more closely the issues of coverage and classification of race and ethnic groups. One issue to be investigated is the issue of context, which can have enormous effects for the self-reporting of race and ethnicity. As noted above, how a question is asked and for what purpose can affect self-identification responses for race and ethnicity and may lead to differences for the census compared to other data. Another issue is one of changes in the social environment: public attitudes can shape how the nation thinks of its race and ethnic groups and how individuals think of themselves. For instance, there have been enormous gains in the number of respondents to the decennial census who report themselves as American Indian. Evidently, rapid shifts have occurred in how individuals think of themselves as American Indians or in their willingness to report this racial identity on the census. This is an important issue for further study.

#### *Federal Needs*

A key need for census data comes from federal agencies. Some agencies have mandated uses of census data, other agencies have mandated needs to produce statistics that rely partly on census data, and other agencies have programmatic needs for census data. Census data are used in several ways: benchmarks for data series, sampling frames, and for stratification and standardization. They are also used for interpretative purposes by federal agencies (i.e., tabulating income by work experience and for calculating such other measures as birth and death rates by age, sex, and race). And they are used for special interest tabulations by other federal agencies, either on a cost-reimbursable basis or special request.

The panel will consider the characteristics of a census that includes data to fulfill federal program or research purposes. To do so, the panel will need to examine what information is available on mandated requirements for census data. The panel will review the effects of Title 13 on assembling microdata for state and local needs. Many federal mandates require states and cities to use census data for their program administration. Questions therefore arise about the usefulness of having increased state and local access to databases that would allow them to assemble microdata, using federal and their own information. The panel will also investigate the use of census data when it is not mandated but is routinely used.

#### *State and Local Needs*

One of the most important needs for census data comes from users of small-area data for states and local areas. There are several types of data users: state and local governments, the research community, private business, and non-profit groups. In its work for the final report, the panel will consider the mandatory and non-mandatory needs for item content in the census, including the criteria (accuracy, completeness, timeliness, comparability, and confidentiality) for alternative census designs.

#### *Administrative Records*

As discussed above the panel believes that it is important to examine the potential of administrative records to complement or perhaps substitute for all or part of the decennial census. A good beginning for the census use of administrative records would be to increase the exploitation of those records for small-area estimates. The expanded use of administrative records for intercensal estimates would provide ex-



perience in working with records, revealing their limitations for content and coverage, noting the quality problems, and demonstrating access and timeliness issues.

The panel plans to look at the problems associated with the use of administrative records for the decennial census. There are some known problems with administrative records: they do not cover the entire population, they have limited content, there is a need to avoid duplication if multiple lists are merged, and the quality of the records may vary. The panel will also examine issues of legal authority and interagency collaboration, including questions of access to records and the potential for records being linked to individual or household data. The panel intends to look at Title 13 issues for the use of administrative records by the Bureau of the Census.

The panel believes that it is important to assess the value of maintaining a continuous inventory of housing referenced to correct geographic location. Such an inventory would be valuable administrative data in itself as well as a database for geographically referencing other administrative records to their correct geographic location and for linking administrative records to the correct household.

As evidenced in its recommendation (above), the panel believes that work on administrative records needs to begin soon. Over the past decade, research on administrative records has not been aggressively developed, and there is little formal incorporation of administrative records in the decennial census program. The panel notes again that there are three potential uses of administrative records: as an alternative to the census, for census coverage improvement, and for intercensal estimates. Long-term research is needed if knowledge is to be gained about the possible use of redesigned administrative records as a substitute for the traditional census. And research is needed in the near future to assess the potential of administrative records for coverage improvement for the 2000 census. The panel plans to examine whether there is a role for administrative records in the census and, if so, how much of a role they might play.

#### *Sampling for Nonresponse*

Sampling for the follow-up of nonrespondents has the possibility of big cost savings for the decennial census. However, it also requires much more testing. Without methodological testing, information is lacking about cost savings, effects on small-area data, and operational considerations. Sampling may save money, but will it result in negative effects for the quality of small-area data? The panel plans to examine further the potential of sampling for nonresponse, partly in conjunction with methodological considerations of sampling by the Panel to Evaluate Alternative Census Methods.

#### APPENDIX A—DATA REQUIREMENTS FOR REAPPORTIONMENT AND REDISTRICTING

One of the first tasks of the panel was to examine the requirements for the decennial census to satisfy data needs for reapportionment of the U.S. Congress and redrawing of congressional and state and local districts for purposes of political representation. The Constitution of the United States mandates in article I, section 2, that "representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective numbers. . . . The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten years, in such Manner as they shall by Law direct." Hence, the requirement to support congressional reapportionment, which, in turn, entails redrawing congressional district boundaries, represents the absolute bedrock upon which the U.S. decennial census rests.<sup>1</sup>

The panel reviewed data requirements for reapportionment and redistricting (at the federal, state, and local levels) that stem from the Constitution (including amendments) and also from other statutes (e.g., the Voting Rights Act, Title 13 of the U.S. Code), judicial interpretation, and administrative practice. The panel sought to determine how open to interpretation the requirements might be, so that, in turn, it could consider the fullest possible range of census designs in the spirit of a "zero-based" assessment of the most cost-effective ways to conduct future censuses. Specifically, the panel considered the possibility that designs that made use of sampling and administrative records, which might offer cost savings and other bene-

<sup>1</sup> The provision in article 1, section 2 (also in article 1, section 9) that required direct taxes to be based on the census was effectively repealed by the Sixteenth Amendment, ratified in 1913, which stated that "the Congress shall have power to lay and collect taxes on incomes, from whatever source derived, without apportionment among the several States, and without regard to any census or enumeration."

fits for the census, could satisfy data requirements for reapportionment and redistricting.

#### REAPPORTIONMENT

As noted above, the U.S. Constitution mandates the conduct of an "enumeration" every 10 years for the purpose of reapportionment of the U.S. House of Representatives.<sup>2</sup> The interpretation of the word "enumeration" is obviously key to an assessment of whether census designs that involve sampling or administrative records could serve this fundamental purpose. A second important consideration is that reapportionment must be effected simultaneously for the entire country—one cannot reapportion in some areas in one year and in other areas in another.

#### ROLE OF SAMPLING

A legal review prepared by the Congressional Research Service (Lee, 1993) concludes that, for the purpose of reapportionment, there needs to be an attempt to account for every inhabitant in the country: a sample census, no matter how large, cannot satisfy the constitutional requirement. Similarly, "rolling census" designs, in which different parts of the population are surveyed each year without even a minimal census of the entire population at any one time, would not satisfy the requirement. (The rolling census designs proposed by Horvitz [1986] and Kish [1981, 1990] are in this category.) Other rolling census designs, namely, those that do include a minimal census every tenth year, would satisfy the constitutional requirement. (The designs proposed by Herriot et al. [1989] and Alexander [1993] are of this type; see the body of the report for further discussion of the potential and problems of rolling census or continuous measurement designs that include a minimal decennial census together with rolling surveys.)

Lee (1993) draws the conclusion that an attempt at a complete count is constitutionally required from a review of the meaning of the word "enumeration" at the time the Constitution was adopted and subsequent legislative and judicial history.<sup>3</sup> In particular, two sections of Title 13 of the U.S. Code (which pertains to the Census Bureau) address the topic of sampling in the decennial census. Section 195, adopted in 1957, states that except [emphasis added] for the determination of population for purposes of apportionment of Representatives in Congress among the several States, the Secretary shall, if he considers it feasible, authorize the use of the statistical method known as 'sampling' in carrying out the provisions of this title." Section 141(a) appears to be more liberal, in that it authorizes the Secretary of Commerce to take a decennial census every 10 years "in such form and content as he may determine, including the use of sampling procedures and special surveys." Lee (1993) states that the courts have reconciled the two sections by holding that the Census Bureau may use sampling procedures in the census but only in addition to more traditional methods of enumeration: see, for example, *Carey v. Klutznick* (1980, Southern District, New York); other relevant cases are referenced below.

Although a sample census and some rolling census designs appear precluded on constitutional grounds, the use of sampling as part of the census process appears compatible with the spirit of the constitutional, legislative, and judicial history regarding "enumeration," so long as the process includes an effort to reach all inhabitants. Specifically, designs that use sampling for the follow-up stage of census operations (after an initial attempt has been made to deliver a questionnaire to every household or person) and for coverage improvement programs (including adjustments based on sample surveys) would appear to meet the data requirements for reapportionment.

Several court cases (in addition to *Carey v. Klutznick*) have explicitly upheld the constitutionality of an adjustment, citing the importance of having as accurate data as possible for reapportionment and redistricting (see Lee, 1993:18-20; NCSL Reap-

<sup>2</sup> The current reapportionment formula, which uses the method of "equal proportions," was written into law at the time of the 1940 census (Anderson, 1988:189).

<sup>3</sup> Lee (1993:1) notes that the wording changed from "census" to "enumeration" during the course of the Constitutional Convention. She could find no reason for the change but speculates that the negative references to censuses in the Bible may have been a reason (see also Anderson, 1988:10). However, the phrase "census or enumeration" appears in article I, section 9, of the Constitution, which deals with the levying of direct taxes. Lee (1993:2) notes that, in the dictionaries of the time, "enumerate" had the meaning of "to reckon up singly," and "enumeration" had the meaning of "the act of counting over."

portionment Task Force, 1989:4-5).<sup>4</sup> In *City of New York v. U.S. Department of Commerce* (1990, Eastern District, New York), the court stated that it "is no longer novel or, in any sense, new law to declare that statistical adjustment of the decennial census is both legal and constitutional. This Court has already recognized that Article I, Section 2 require(s) that the census be as accurate as practicable." Most recently, the Eastern District Court in New York reaffirmed the conclusion that "the Constitution is not a bar to statistical adjustment" (*City of New York v. U.S. Department of Commerce*, 1993). Although the court upheld the constitutionality of adjustment, it ruled that the decision of the Secretary of Commerce not to adjust the 1990 census results followed the guidelines developed by the Commerce Department for the adjustment decision and could not be deemed to be arbitrary or capricious.

The question of the legality of sampling for nonresponse follow-up has never been explicitly raised in the courts; however, language used in the court cases just cited clearly seem to support its use. For example, in *Carey v. Klutznick* (1980, Southern District, New York), the court held that the Census Bureau may use sampling procedures in addition to a traditional enumeration. In *Young v. Klutznick* (1980, Eastern District, Michigan), the court noted that, since 1970, the census has not been a "simple straight forward headcount" but instead "a relatively accurate estimate of the population developed through the use of self-enumeration by questionnaire, statistical techniques, and computer control devices." The court held that section 195 of Title 13 did not prohibit the use of statistical techniques in the census. (As Lee [1993:20] observed, the court appeared to treat "statistical techniques" as equivalent to "sampling.")<sup>5</sup>

There are precedents for designs that use sampling in the later stages of the census process. In the 1970 census, two coverage improvement programs were conducted on a sample basis, and the results were used to add people to the census by an imputation procedure.<sup>6</sup> In the 1990 census, a postenumeration survey was conducted of a sample of housing units for purposes of evaluating the completeness of the population count and developing adjusted counts on the basis of the sample survey results. These adjusted counts were not used for reapportionment, but the court-ordered process under which they were developed certainly contemplated that they might be so used (see *City of New York v. U.S. Department of Commerce*, 1989, 1990, Eastern District, New York).

#### ROLE OF ADMINISTRATIVE RECORDS

The use of administrative records (e.g., income tax or social security records) to provide complete population data for reapportionment raises a different set of issues from the use of sampling. The use of records is probably not consistent with most people's idea of a census, in that there would be no attempt to contact all the people on or close to a designated "census day." However, such use could be viewed as meeting the constitutional requirements for reapportionment, if there were an administrative records system (or a combination of systems) that, when used for purposes of a census, could be determined to contain data for all inhabitants (or as close to the total of all inhabitants as has been achieved for traditional U.S. censuses), with the records assigned to the correct state of residence. See the body of the report for discussion of both the potentials and the problems of exploiting administrative records for the U.S. census.

<sup>4</sup> These cases include *Young v. Klutznick* (1980, Eastern District, Michigan), *City of Philadelphia v. Klutznick* (1980), and *City of New York v. U.S. Department of Commerce* (1990, Eastern District, New York). *Cuomo v. Baldrige* (1987, Southern District, New York) also implicitly assumes that an adjustment would be constitutional.

<sup>5</sup> In work commissioned by the panel, Barnett (1993) and Issacharoff (1993) agree with Lee (1993) that court decisions to date allow for the use of statistical methods, including sampling, as part of the census process. They go even further to offer arguments whereby the Constitution, which allows Congress complete discretion to determine the method for taking the census, might be interpreted to permit the use of methods that do not involve any physical enumeration at all. In their view, the key legal requirement is that the population figures be obtained by the most reliable methods possible.

<sup>6</sup> The 1970 National Vacancy Check involved a resurvey of 13,500 housing units originally classified as vacant (about 0.2 percent of all such units). On the basis of the findings, imputation procedures were used to reclassify 8.5 percent of all vacant units as occupied and to impute persons to these units, amounting to 0.5 percent of the total population count. The Post-Enumeration Post Office Check was conducted in rural areas of 16 southern states. The Postal Service checked the address lists developed by enumerators for completeness, and Census Bureau staff followed up a sample of missed addresses in the field. On the basis of the results, census records were imputed for 1.3 percent of the population in these areas, representing 0.2 percent of total U.S. population; see Citro and Cohen (1985:189-193).

There is no body of legal opinion on which to base an assessment of the constitutionality of an administrative records census. However, both Lee (1993:30) and Barnett (1993) conclude that a census based on administrative records would likely be held by the courts to be constitutional if it could be demonstrated that the data were accurate.

With regard to the more limited use of administrative records as part of the U.S. census process, there are several precedents. The Nonhousehold Sources Program was an administrative records-based operation to improve coverage in the 1980 census (Citro and Cohen, 1985:94,200). It involved matching several lists to census records for selected census tracts in urban district offices. The lists used were driver's license records, immigration records, and public assistance records in New York City. Enumerators visited addresses of people identified from the match who might have been omitted from the count. The Parolee/Probationer Check was an administrative records-based coverage improvement program adopted in the 1990 census (Ericksen et al., 1991:43-47). As part of this operation, probation officers in large cities and smaller cities with large minority populations were asked by census enumerators to verify addresses of parolees and probationers obtained from records. These addresses were matched against the census, and all cases of nonmatches were added to the census, with no attempt at a personal follow-up. In other words, this program added individuals to the census based solely on administrative records, as might be done for an entirely administrative records-based census.

Both programs had serious problems of implementation that indicate needed areas for further research and development to improve the data quality and cost-effectiveness of coverage improvement efforts that make use of administrative records. The 1980 Nonhousehold Sources Program had very low payoff in terms of additions to the census count. A total of about 6.8 million records were checked against the census, but only 130,000 people were added to the count as a result of the matching and field follow-up operations (Citro and Cohen, 1985:200). The Parolee/Probationer Check had a high error rate. An estimated 53 percent of the additions to the census count as a result of this program (about 250,000 people) were erroneous enumerations, that is, people who were already counted or who should not have been included for some other reason (Ericksen et al., 1991:44).

## REDISTRICTING ON THE BASIS OF TOTAL POPULATION

### LEGAL FOUNDATIONS

Reapportionment of the U.S. Congress or a state or local legislature carries the implication that district boundaries should be redrawn to accommodate changes in the number of seats allotted to the jurisdiction and, even if that number does not change, to accommodate changes in the distribution of population so that a vote in one district carries about the same weight as a vote in another. Indeed, in the nineteenth century, Congress typically passed a statute at the time of each census that required all states, whether or not they gained or lost seats, to redistrict and to establish single-member districts that were contiguous, compact, and as nearly equal in population as practicable (Durbin and Whitaker, 1991:4-5).<sup>7</sup> After the 1920 census, however, Congress declined to reapportion the House because of concerns of rural interests about the tremendous population growth in the cities, particularly from immigration. In 1929, Congress passed an act that provided for automatic reapportionment upon delivery of the population counts after each census, but it set no standards for redistricting. The courts held that the omission of such standards was intentional, and, since then, it has been up to the courts themselves to provide for any standards (Durbin and Whitaker, 1991:4-5).<sup>8</sup>

From the 1920s through the 1950s, the courts generally declined to intervene in the "political thicket" of redistricting, and congressional and state legislative districts became increasingly more unequal in population size. Many states chose not to redistrict after a census, unless they gained or lost seats, and those that did often paid little attention to achieving population equality across districts. Very large deviations in population, generally favoring rural over urban and suburban districts, were quite common. After the 1960 round of reapportionment and redistricting, the largest congressional district in the U.S. had over five times the population of the

<sup>7</sup> Until 1911, with one exception, Congress increased the size of the House of Representatives at each reapportionment, so that no state lost congressional seats. The exception occurred when the size of the House was decreased from 242 members following the 1830 census to 232 members following the 1840 census. In 1911, the size of the House was fixed at 435 seats.

<sup>8</sup> See Anderson (1988:Ch.6) for a review of reapportionment and redistricting history and the relationship to the census.



smallest district; the 20 most populous districts had a combined population of 14 million compared with a combined population of 4.6 million for the 20 smallest districts. Disparities among state legislative districts were even greater (Baker, 1986:258).

A 1967 law required single-member congressional districts. Also, the Voting Rights Act, as interpreted by the courts and administrative practice, led to de facto standards with regard to the representation of minorities. (The Voting Rights Act was enacted in 1965 and has been extended and amended several times since then.)

The landmark "one-person, one-vote" Supreme Court decisions, beginning in the early 1960s, changed the requirements for redistricting drastically. In the first of these cases, *Baker v. Carr* (1962), which involved Tennessee state legislative districts, the court held that reapportionment and redistricting matters were subject to judicial review under the equal protection clause of the Fourteenth Amendment. In *Wesberry v. Sanders* (1964), the court held, under Article 1 of the Constitution, that congressional districts must be as nearly equal in population as practicable. In *White v. Weiser* (1973), the court rejected a Texas congressional redistricting plan in which the smallest district was about 4 percent smaller than the largest district, and in *Karcher v. Daggett* (1983), the court in a 5-4 decision rejected a New Jersey congressional redistricting plan in which the smallest district was only 0.7 percent smaller than the largest district. The court held that the state could have avoided such a deviation, as it had rejected a plan with a population deviation of only 0.45 percent. Furthermore, the court ruled, the state had failed to show that the deviation in its approved plan was needed to achieve a legitimate goal (Parker, 1989:61; see also Durbin and Whitaker, 1991:12; and Ehrenhalt, 1983:56-57).

Whether the Supreme Court will continue to view virtually absolute population equality among congressional districts as an overriding constitutional requirement is, of course, not certain. Although *Karcher v. Daggett* (1983) is certainly indicative of a strict interpretation, the decision was a close one (5 to 4), and language in both the concurring and dissenting opinions at least raises the possibility that the court could modify its pursuit of absolute population equality in the future.<sup>9</sup>

Over the same period, the Supreme Court issued decisions that greatly affected state and local as well as congressional redistricting. In *Reynolds v. Sims* (1964), the court held that, under the Fourteenth Amendment, both houses of a state legislature must be apportioned on a population basis.<sup>10</sup> Moreover, although mathematical exactness may be impossible, states should strive for population equality. Generally, however, the courts have allowed more deviation among state legislative seats than among congressional districts. The guidelines appear to be that deviations of up to 10 percent in the size of state districts are constitutionally acceptable, although they can be challenged on other grounds (e.g., racial discrimination). Deviations between 10 and 16 percent are presumed to be unconstitutional, but states can try to justify them; deviations above 16 percent are usually viewed as completely unacceptable (Parker, 1989:57-58; see also O'Rourke, 1980:22).

#### CENSUS DATA FOR REDISTRICTING

After the 1970 census, the states could obtain population counts for geographic areas as small as city blocks, which were defined in urbanized areas and in other localities that contracted with the Census Bureau, and for enumeration districts in unblocked areas. However, no special data files or reports were provided specifically to meet redistricting needs. In 1975 Congress required the Census Bureau to provide decennial census population tabulations to state officials for purposes of legislative reapportionment or redistricting within one year after the census date (i.e., under the current schedule, by April 1 of census year plus one) (Public Law 94-71; section 141(c) of Title 13). States can specify the geographic areas for which they require

<sup>9</sup> One commentator (Baker, 1986:275-276) claims that a majority of the Supreme Court no longer truly supports the ideal of strict mathematical equality for congressional districts but has felt constrained by precedent. He argues that Congress should pass legislation that would permit a reasonable degree of population variance among districts and require other desirable criteria, such as compactness and contiguity.

<sup>10</sup> This decision effectively abolished systems that assigned a minimum number of seats to each county or other jurisdiction no matter how small in size, although one exception occurred for a Wyoming state legislative redistricting plan. On the same day as the Supreme Court issued *Karcher v. Daggett* (1983), it issued *Brown v. Thomson* (also a 5 to 4 decision), which upheld a plan that gave a seat to one county that was significantly smaller in size than the other districts. The court accepted the state's argument that this deviation from the ideal was justified to permit one isolated county to retain the seat it had been granted in 1913. Two members of the majority issued an opinion saying that they agreed with the decision only because the deviation from population equality involved just one county (Parker, 1989:61).

tabulations, provided that their requirements satisfy Census Bureau criteria and are transmitted to the Bureau no later than 3 years prior to the census date; if no special areas are identified, the Census Bureau is to provide "basic tabulations of population."

In practice, "basic tabulations" has come to mean tabulations for individual blocks, the smallest area of geography identified in census data products. After the 1980 census, the P.L. 94-171 data file provided by the Census Bureau contained the following information: total population; white; black; American Indian, Eskimo, and Aleut; Asian and Pacific Islander; and other races; and a separate count of Hispanics. Data were provided for states, counties, minor civil divisions, places, voting districts (where specified by the state), census tracts or block numbering areas, enumeration districts or block groups, and blocks. The number of blocks identified in the 1980 census was 2.5 million, an increase from 1.7 million in 1970. Blocks were identified in all urbanized areas, all incorporated places of 10,000 or more population, and other areas for which a state or local government contracted with the Census Bureau to define block boundaries. The average population per block was estimated in the 1980 census at about 70 people, and the average population per enumeration district at about 600 people (Bureau of the Census, 1982:56,67,79). Voting precincts identified by the states were generally the size of an enumeration district or group of blocks.

In 1990 the P.L. 94-171 file was expanded to include, in addition to the 1980 content, cross-tabulations of all items by age (under 18 and 18 years and over), a cross-tabulation of race by Hispanic origin, and a count of occupied housing units (included in response to requests from other users).<sup>11</sup> The geographic areas identified in 1990 were the same as in 1980, except that enumeration districts as a concept no longer existed and blocks—about 10 million in all—were defined for the entire country (Bureau of the Census, 1992a:82-83).

Although P.L. 94-171 requires the Census Bureau to furnish decennial census tabulations to the states, the courts have clearly held that the states may use other data sources for redistricting purposes. Over time, however, the states, on their own initiative and prodded by the courts, have come to rely almost exclusively on census data to prepare redistricting plans.

In the 1960s and 1970s, several court cases held that a state could use other than decennial census data for congressional and other kinds of legislative districts (NCSL Reapportionment Task Force, 1989:12-13). In 1966, citing Hawaii's special military and tourist populations, the Supreme Court in *Burns v. Richardson* held that the state could redistrict on the basis of numbers of registered voters. However, this decision was reached after the court determined that the results would not have been substantially different from those based on total citizen population. In *Kirkpatrick v. Preisler* (1969), the court implied that the eligible voter population could be the basis for redistricting if identified properly and applied uniformly. In *Ely v. Klahr* (1971), the court cautioned that a new plan for Arizona legislative districts could use registered voter data only if the results would not differ substantially from what would have resulted "from the use of a permissible population base."

In the 1980s, the case law generally reflected the position that alternative sources of data could be used if they were applied uniformly and the results were comparable to those under a plan based on the total or total citizen population. For example, the District Court in Massachusetts upheld the use of a state census for legislative redistricting in *McGovern v. Connolly* (1986). However, a district court struck down a New Mexico plan that was based on number of votes cast, and, in 1982, the District Court in Hawaii struck down Hawaii's state legislative plan that used registered voters, finding that the results did not "substantially approximate" those based on total population. In the same case (*Travis v. King*, 1982), the District Court struck down Hawaii's congressional redistricting plan, also based on registered voters, as unconstitutional: "(P)ursuant to Article I, Sec. 2 of the Constitution states must depend on total federal census figures to apportion congressional districts within their boundaries."

Recent cases dealing with the issue of adjustment of the census for coverage errors have generally upheld the view that the states need not use census data, at least not exclusively (Lee, 1993:5-5). For example, in reviewing *Young v. Klutznick*, the appeals court (1981, 6th Circuit) held that states are not constitutionally required to use census data supplied by the Census Bureau for redistricting, but can use adjusted population figures, so long as the adjustment was thoroughly docu-

<sup>11</sup> The additional items provided in the 1990 P.L. 94-171 file were available from the 1980 census as well but only in data files that were released at a later date than the 1980 P.L. 94-171 file; hence, they were not available for redistricting on a timely basis.

mented and systematically applied. In *City of Detroit v. Franklin* (1992, Eastern District, Michigan), the court noted that *Karcher v. Daggett* did not hold that states must use census figures in redistricting but rather must use "the best population data available." In Senate of the *State of California v. Mosbacher*, the appeals court (1992, 9th Circuit) held that the Census Bureau was under no obligation to release adjusted data, but, if the state knew the census data were underrepresentative of the population, it could and should use noncensus data, in addition to the official count, for redistricting. In the most recent case (*City of Hew York v. U.S. Department of Commerce*, 1993, Eastern District, New York), the court found that there was a public interest in having available data tapes containing adjusted 1990 census counts down to the block level for the entire United States. The court ruled that the plaintiffs, which had acquired these tapes from the Census Bureau as part of the court-ordered process for deciding whether to adjust the census, could make the tapes publicly available.<sup>12</sup>

Over time, virtually all states have come to rely on census population counts for legislative redistricting. When states have used other bases, such as registered voters, they have generally had to obtain census data to demonstrate to the courts that their data would not give a substantially different result from census data. Although the court cases on adjustment noted above appear to give considerable discretion to the states in their choice of population data, they deal largely with the question of whether to use the official census counts or adjusted figure that are also based on the census. It seems clear that population counts for redistricting are a practical, if not precisely a constitutional, required use of the U.S. census.

#### REQUIRED LEVEL OF GEOGRAPHIC DETAIL

Over time, the need for block data from the census for redistricting has also become a practical requirement: although not mandated, block data are treated by most parties as if they were mandatory. The driving force behind this focus on block data appears to come from the stipulations of the courts that there be virtually no deviation among congressional districts in population size and very little deviation among state legislative districts.

In fact, not all states actually use block data in the redistricting process; many use data for election precincts or voting districts for which they have specified the boundaries to the Census Bureau (under the provisions of P.L. 94-171). These districts are generally the size of a block group, or enumeration district in the old terminology. Commenting on the 1980 census P.L. 94-171 data program, officials of 33 states said they preferred working with block totals, and 15 states indicated they preferred working with voting district summaries. A total of 28 states, however, wanted to see voting district summaries for the entire state and 8 states wanted to see such summaries for portions of the state in order to relate voting data to the block or precinct population data (Romig, 1983:9-10; see also Bureau of the Census, 1983b; for a preliminary assessment of reactions to the 1990 P.L. 94-171 program, see National Conference of State Legislatures, 1992). The Census Bureau requires that voting district boundaries follow streets or other geographic features so that they align with census geographic boundaries; however, the precincts need not and typically do not align with geographic levels larger than the block (e.g., they often cut census tract boundaries). Hence, the Census Bureau has used block data to provide voting district summaries for the states that request them.

As noted above, the courts have given primacy to the population equality of districts over such other criteria as compactness or contiguity. The result has often been the creation of very peculiar-looking districts, as state legislative majorities seek to redraw district boundaries in a manner that maximizes partisan advantage.<sup>13</sup> Indeed, another force to retain blocks as the basic unit for redistricting is the widespread use of voting data (e.g., percent Democratic and Republican) in conjunction with census data to determine district boundaries. Typically, the people who work on redistricting want to look at several years of voting data because precinct boundaries change frequently: a common practice is to allocate precinct voting data to blocks and reaggregate the blocks to try out various redistricting plans to determine their political advantage.

<sup>12</sup> Previously, the Department of Commerce had released adjusted census data for the nation, states, counties, and cities; it had also released to the Congress adjusted data for half of the census blocks.

<sup>13</sup> In *Davis v. Bandemer* (1986), the Supreme Court held that partisan political gerrymandering was subject to judicial review, but it set a high standard for successfully bringing such a case, stating that plaintiffs must do more than show that a redistricting plan makes winning elections more difficult (Durbin and Whitaker, 1991:13-14).

All of this manipulation of small-area data has been made possible by computers and the advent of the Census Bureau's TIGER geocoding system. Indeed, one firm that assisted about half the states with redistricting after the 1990 census expressed the wish that census data could be provided for block faces (i.e., sides of blocks). The widespread use of computers as a data processing and mapping tool has made it easy for legislators to examine a variety of plans in order to determine the best one from their point of view. In turn, this behavior has made it necessary to have a database that can be reaggregated in many ways.<sup>14</sup>

#### ROLE OF SAMPLING AND ADMINISTRATIVE RECORDS

It seems clear that the provision of population figures by block is, for all practical purposes, a burden that the U.S. census must satisfy. Even if the restrictions on allowable deviations in population size were to be relaxed somewhat in the future, small-area data would still be required to provide the basic units for definition of legislative districts. This requirement all but eliminates the "sample census" design, even if such a design were otherwise determined to be constitutional and feasible. The sample size would have to be so large it would practically preclude any cost savings compared with a complete count census.

There appears to be no constitutional barrier to the use of sampling as part of the census process and hence no reason to rule out its use *a priori*, but questions have been raised about the quality of block-level data that might result under designs that incorporate sampling. For example, a factor in the decision not to use adjusted census counts in 1990 or, more recently, for intercensal estimates, was the belief that the adjustment factors developed on the basis of the postenumeration survey were not sufficiently reliable for small geographic areas (Bryant, 1993). Certainly, the merits of sampling as part of the census process require thorough research to determine its effect on total error and the costs of implementing various types of sample operations. It is possible, however, that the net effects could be positive: for example, careful sampling for nonresponse follow-up, at an appropriate sampling rate, might reduce costs and also reduce total error by decreasing nonsampling error more than the added variability due to sampling. Moreover, it is critical to keep in mind that block data are the input to redistricting, not the output. So long as the block data are of sufficient quality that, when aggregated to the congressional or other district level, the quality of the estimated population of the larger area is high, then the data requirements for redistricting would be served.

Census designs that make use of administrative records might also meet the practical requirement for population totals at the block level for redistricting purposes, if a number of problems can be overcome. One problem for the purpose of redistricting—which is likely to be much more severe than the corresponding problem for congressional reapportionment—concerns the accuracy of the addresses in administrative records systems, which may be out of date or for a business or for another individual (e.g., a tax preparer).

#### REDISTRICTING ON THE BASIS OF RACIAL OR ETHNIC COMPOSITION

All of the above discussion of data needs for reapportionment and redistricting has focused on total population figures. The civil rights movement of the 1950s and 1960s led to legislation, court decisions, and administrative practices that moved another requirement front and center, namely, the need for data on race and ethnic origin for purposes of legislative redistricting.<sup>15</sup>

#### LEGAL HISTORY

The Voting Rights Act, originally passed in 1965 (P.L. 89-110) and extended and amended in 1970, 1975, 1982, and 1992, has led to the practical necessity for race

<sup>14</sup> This discussion of redistricting practices draws heavily on information provided to the panel staff by Kimball Brace, president of Election Data Services, Inc., Washington, D.C. See also Ehrenhalt (1983) for an overview of the redistricting experience following the 1980 census; and Henry (1989), who notes that half the states used computers for tabulation purposes in the 1980 redistricting cycle and that (as of 1989) several firms were ready with integrated computerized tabulating and mapping systems for the 1990 cycle.

<sup>15</sup> Ethnic origin in the context of redistricting generally refers to Spanish or Hispanic origin. In addition to total population and information on race and ethnicity, many states want data on the voting age population. These data were added to the P.L. 94-171 data file in 1990, and 28 states reported using them (National Conference of State Legislatures, 1992:8). The requirements for the census as the source of such data are not considered further here. It can well be argued that age—and sex—are essential characteristics to obtain simply to establish the existence of an individual and to help ensure that people are not double-counted.



and ethnic data from the census (see Laney, 1992, for a history of the act). The act nowhere actually stipulates the use of census data, although it does require the director of the Census Bureau to make a number of determinations. But interpretations of the act by the courts and the Justice Department have virtually mandated the need for census data in redistricting.

The original intention of the Voting Rights Act was to make it possible for blacks in the South to obtain the opportunity to participate in elections, an opportunity which was often denied them by unreasonable literacy tests and other barriers to registration and voting. The act, as enacted in 1965, prohibited (in section 2) under the authority of the Fifteenth Amendment the enactment of any election law to deny or abridge voting rights on account of race or color. It further specified (in section 4) that any state or county that had any test or device as a condition for voter registration on November 1, 1964, and in which the number of registered or actual voters fell below 50 percent of total voting age population in the 1964 presidential election could not use a literacy test or any other test or device to screen potential voters. Finally, it provided (in section 5) that any covered jurisdiction (i.e., any jurisdiction required to drop voting tests under section 4) had to submit "any voting qualification or prerequisite to voting, or standard, practice, or procedure with respect to voting" adopted after November 1, 1964, for "preclearance" to the U.S. Department of Justice or the U.S. District Court for the District of Columbia to determine that there was no abridgement of the right to vote on the basis of race or color.

The 1970 amendments to the Voting Rights Act outlawed literacy tests and other devices in all jurisdictions, not just those in covered jurisdictions, and extended coverage to jurisdictions that had such tests in November 1, 1968, and in which there was less than 50 percent registration or turnout in the 1968 presidential election. The effect of this provision was to cover subdivisions in northern and western as well as southern states.

The 1975 amendments to the act included a major new provision that, on the basis of the equal protection clause of the Fourteenth Amendment, extended coverage under the act to protect the voting rights of "language minorities," defined to be people of Spanish heritage, American Indians, Asian Americans, and Alaskan Natives. The preclearance provisions of the act (i.e., the requirement to submit proposed changes in voting procedures to the Justice Department for approval) were applied to any jurisdiction (counties and independent cities in most states and townships in others) for which the Census Bureau determined that more than 5 percent of the voting age citizens were of a single-language minority, election materials had been printed only in English for the November 1972 elections, and less than 50 percent of all voting age citizens in the jurisdiction had registered or voted in the 1972 presidential election. This provision covered the states of Alaska, Arizona, and Texas and political subdivisions in eight other states.

The 1982 amendments to the act kept the basic provisions intact, but made some changes. The amendments extended the preclearance section of the act (likewise the provision for examiners and election observers in covered jurisdictions) to the year 2007, but also provided that Congress reexamine them in 1997. (Jurisdictions can petition for release from the provisions at an earlier date, but they must meet a stiff set of criteria for release.) Another provision stated that the standard of proof for judging an election law to be discriminatory under section 2 (as well as section 5) was no longer discriminatory intent, but, rather, discriminatory result. As somewhat of a counterbalance, still another provision stated that minorities had no right to proportional representation, but that the courts could consider the lack of representation as part of the totality of circumstances in cases brought under the Voting Rights Act.

Finally, although not related to redistricting data needs per se, the act, as amended, included a provision (section 203) that is currently satisfied by using data from the census long form. This provision, as first adopted in 1975, required jurisdictions (counties, cities, or townships) to provide election materials and oral assistance in another language as well as English in areas for which the Census Bureau determined that 5 percent of the voting age citizens were of a single-language minority and the illiteracy rate in English of the minority (defined as failure to complete 5th grade) was greater than the illiteracy rate in English of the entire nation. The 1982 act amendments asked the Census Bureau to investigate the usefulness of 1980 census long-form questions on "mother tongue" and English-speaking ability for determining coverage under bilingual assistance provisions. On the basis of the Census Bureau's research, the definition of a covered area became one in which 5 percent of the citizens of voting age comprised a group of people who spoke a single minority language and who said they did not speak English very well and who had a higher

illiteracy rate than the nation as a whole. The result of this change was to reduce the number of covered areas from about 500 following the 1970 census to about 200 following the 1980 census; about 300 covered areas were identified after the 1990 census. The 1992 amendments extended the bilingual voting assistance provisions to 2007 and made some additional minor changes to the definition (see Bureau of the Census, 1976a; Kominski, 1985, 1992).

With regard to data needs for redistricting, it is section 5 of the Voting Rights Act, with its requirement for federal review and preclearance—not only of tests for voting registration but of any “standard, practice, or procedure with respect to voting—that has led to the practical necessity for census figures on race and ethnicity by block for redistricting.<sup>16</sup> A key case that supported the use of section 5 to review many aspects of states’ (and counties’ and cities’) electoral systems was *Allan v. Board of Elections* (1969), in which the Supreme Court held that such changes as moving from single-member to multimember districts were “practices or procedures” that were subject to review under section 5 because they had the potential of “diluting” the black vote. The Justice Department quickly moved to instruct legal officers in covered jurisdictions to clear every change in voting procedure. Whereas only 323 voting changes were received by the Department for preclearance between 1965 and 1969, almost 5,000 were submitted between 1969 and 1975 (Thernstrom, 1979:59).<sup>17</sup>

From 1965 to 1988, the Justice Department most often objected to three ways of setting up electoral systems on the grounds that they would have resulted in abridging or diluting the voting power of blacks, Hispanics, or other protected minority voters (Parker, 1989:Table 6.1): municipal annexations (1,088 of 2,167 total objections over the period); changing from single-member districts to at-large voting (472 objections); and redistricting plans that lessened the effectiveness of minority votes, for example, such schemes as dividing concentrations of minority voters into adjoining majority white areas or minimizing the number of minority districts by placing minority voters in as few districts as possible (248 objections). Congress also specifically expressed an interest in having redistricting plans reviewed by providing in the 1975 amendments that jurisdictions covered as of 1965 could not seek release from coverage until 1982 and by extending that date in the 1982 amendments to 1984. These dates were enacted to ensure that there would be time for Justice Department or court review of redistricting plans based on the 1980 census in those jurisdictions (Laney, 1992:18, 24).

#### ROLE OF THE CENSUS

As noted above, the data files provided to the states by the Census Bureau (under P.L. 94-171) include race and ethnic counts by blocks—and voting districts if specified by the state—in addition to total population. The 1990 file added a cross-tabulation of race by Hispanic origin. In a survey of reactions to the 1990 redistricting data program, 32 states said they used the race and ethnic data, and 3 said positively that they did not (National Conference of State Legislatures, 1992:8). As noted above, not all states are covered by the preclearance provisions of the Voting Rights Act, although, potentially, redistricting plans in any state can be challenged under section 2. Overall, it seems clear that, for many states, the data are an important input to the redistricting process.

Whether this situation will continue unchanged in the future is a question. The preclearance procedures of the Voting Rights Act are scheduled to be reviewed by Congress in 1997 and, in any event, to expire in 2007. However, the language of the 1982 amendments, although not guaranteeing the right to proportional representation, states that minority representation is a factor to be considered by the courts. Also, historically, the preclearance provisions have been extended every time the act has come up for renewal. It would seem prudent to expect that the census in the

<sup>16</sup> Parker (1989:59-63) notes that challenges to redistricting plans on the grounds that they are racially discriminatory can be brought under section 2 of the act as well as the more frequently invoked section 5. Hence, although the preclearance provisions of section 5 currently apply to fewer than half the states (in some cases, just to selected jurisdictions in the state; see Laney, 1992), all states must worry about the racial composition of legislative districts if they are to avoid challenges under the Voting Rights Act.

<sup>17</sup> See Davidson (1992), Grofman et al. (1992), Karlan and McCrary (1988), Kousser (1992), and Thernstrom (1979, 1987) for discussions of the history and implications of the Voting Rights Act from different perspectives.

year 2000 and most likely in the year 2010 will need to supply block-level data on race and ethnicity for purposes of legislative redistricting.<sup>18</sup>

Adding a characteristic such as race or ethnicity to census requirements raises the issue of measurement error and, for designs that make use of administrative records, the issue of data availability. Several questions and concerns with regard to obtaining race and ethnicity data from alternative census designs are raised in the text of the report, and further consideration of this topic will be an important part of the panel's work.

## APPENDIX B—NEW SOURCES OF SMALL-AREA DATA

This appendix on new systems for providing small-area data is presented for two reasons. The first is to suggest that there are alternative sources to the traditional census for many types of small-area data. For example, as noted in the body of the report, small-area data derived from administrative records can offer more frequent estimates. The second purpose is to invite comments from readers on the idea of developing new sources of small-area data. The panel does not endorse any of the data systems discussed in this appendix. Rather, we describe some examples of new small-area data systems and their characteristics and uses, noting some of the advantages and limitations. Discussion of changes in the decennial census have usually increased the anxiety of small-area data users by appearing to "threaten" them with the potential loss of data. Heightened anxiety need not arise if two interrelated points are emphasized: that small-area data will be provided by future census activities and that a "zero-based" appraisal of the census requires some consideration of reasonable alternatives to the current design.

We refer to three databases in this appendix: (1) a geographic database composed of a computerized cartographic file and an address file; (2) program data, which would come from available administrative records; and (3) data from the decennial census. A geographic database is a listing of all physical structures and their locations and can be used, by itself, for a continuous housing inventory of small geographic areas. One such geographic database for housing and population data, the TIGER (Topologically Integrated Geographic Encoding and Referencing) system, has already been developed by the Bureau of the Census.

### ADMINISTRATIVE RECORDS AND A GEOGRAPHIC DATABASE

#### ADVANTAGES

Recent census innovations, coupled with technology improvements, have provided data users with a dramatic improvement in the accessibility and detail of small-area census data. Many users now have relevant small-area data available on CD-ROM disks that they can inexpensively manipulate on a microcomputer. Yet the rapid pace of social change, especially for small geographic areas, has whetted the appetite for even more timely information. As alternatives to the census, more timely information could be derived from either surveys or administrative records. However, frequent surveys with small-area estimates would be prohibitively expensive. Therefore, the key source of up-to-date small-area data will probably be administrative records that have been geocoded (given a geographic reference) or linked to other records that are geographically referenced.

Program data from administrative records could be used in three ways and could come from different sources. Federal, state, and local governments all have potentially useful records. Among the key data of interest for small-area statistics are administrative records on education, social services, health, labor, veterans' affairs, housing, and transportation. Major federal administrative records are handled by the Internal Revenue Service, the Social Security Administration, and the National Center for Health Statistics (vital registration data). The large federal data systems have the greatest population coverage and the most uniformity. State and local records are often good for addresses, but they often lack coverage. Furthermore, the comparability of the data in the records often varies. Federal records, therefore, warrant initial consideration.

<sup>18</sup> Some analysts involved with redistricting efforts to comply with the Voting Rights Act have argued that it is also important to provide small-area data on citizenship from the census as part of the P.L. 94-171 data files (e.g., see Gobalet and Lapkoff, 1993). They find that citizenship rates among the Hispanic population vary greatly across small areas. Because voters must be citizens, legislative districts may not adequately represent the potential voting strength of the Hispanic population unless citizenship status is known. To provide data on citizenship in the P.L. 94-171 files would likely require asking for this information on the census short form.

First, program records can be analyzed by geographic area. This analysis would not require a geographic database or linkage to other program data. For this use of administrative records (e.g., school enrollments, hospital admissions, or police arrests), the data are summarized and mapped for the geographic areas available in the record system itself (e.g., counties, places, zip codes, school districts, neighborhoods, census tracts). The limitation of this approach is that geographic area detail is limited to what is on the administrative records themselves. Also, no cross-tabulation of records with other administrative data sets can be done. However, such use of administrative data can begin quickly, with no requirement of a geographic database.

Program data can also be linked to a geographic database to permit descriptions for small areas down to individual blocks or block faces (sides of blocks). This is a second use of program data. For this use, the administrative data would have to have street addresses, and the geographic database must have specific address ranges or an individual address file. The advantage of this approach is that estimates would be available for small areas, such as census blocks, which can then be aggregated to match various administrative boundaries. For example, if data on automobile ownership were linked to a geographic-address database, one could have block-level data available on automobiles by model, year, and other associated information. Data could be aggregated for transportation zones or other geographic areas relevant for transportation planning.

Third, program data can be linked at the household level to provide the capability for preparing cross-tabulation information. By actually linking more than one administrative record to a household, data in one record can be cross-classified by data in another record. Cross-tabulations of this type require a geographic database with linked individual addresses and a record system that can be linked to a specific address. A special problem exists for assigning individuals to households if there are multiple households at the same address. This is a methodological challenge. An administrative records system for household and family data would need to develop methods for distinguishing households and families, with the correct assignment of individuals.

Cross-tabulation is a key aspect for this use of program data. For example, suppose a policy question arose of whether a new social program was reducing poverty levels. The requisite program data would be program participation, poverty status, and individual addresses for the program data. With only geographic data, an analysis could note whether areas with the new program had lower poverty rates, but the association would clearly be a spatial one. With records linked to households, the analysis would be whether individual households participating in the new program had lower poverty rates.

One vision of a future information system for small-area data is a continuous linkage of administrative records to a geographic database that supplements decennial census operations. Improvements in the coverage and quality of administrative records, especially if used widely by state and local governments, could also occur as a system becomes operational and widespread.

This vision of future small-area data rests on two pillars: an improved, up-to-date geographic database and available administrative records. There are some advantages of an administrative records-based population program, a program including a decennial census and intercensal estimates. The main advantage for small-area data users would be timeliness: estimates would be available for the years between the decennial censuses. Many private-sector organizations already have frequent data, primarily through the collection of transactional data that are linked to large household record systems. The household records are, in turn, linked to address files so that the aggregated data can be provided for small geographic areas. Currently, there are big differences among users of small-area data: some users have more frequent data than others. Private, for-profit organizations, in particular, are more likely to have access to frequent data on small geographic areas than other users.

#### CONCERNS

Before administrative records are made available for use in an administrative records-based population program, some general concerns must be addressed. These concerns include confidentiality of the program, availability of information that is currently provided by the decennial census, and the comparability of administrative records systems.

First, there is a concern about the confidentiality of any program that would use administrative records extensively. Confidentiality of the administrative records-based population program needs to be assured, and safeguards for a program have to be developed. Confidentiality concerns would vary, however, with the type of use



of program data. Data that are geocoded and analyzed at the aggregate level have different confidentiality issues than program data that are linked to a geographic database or to other administrative records. We stress that the issue of confidentiality warrants prominent attention.

Second, administrative records do not contain all the information that is currently provided by the decennial census. Moreover, there are serious questions about the quality of data in some administrative records. In the absence of information, data would need to be estimated from available records or by statistical modeling, using a sample of records. The production of small-area data from administrative records would require decennial census data as a benchmark if the purpose of estimates from administrative records is to provide intercensal data that match census information.

Third, many administrative records systems are comparable only within the geographic area of their collection. State highway data, for example, might be comparable for areas within the state but are likely to differ among states. Although it would not limit the use of small-area data within the geographic unit of their collection, it is an important issue for national comparability of statistics.

An important first step in the development of a new census design for small-area data would be a demonstration project; such a project would provide actual field experience. It would allow data users to become familiar with how the new data are prepared and how they might be used, including differences in such use from current data. The provision of small-area income estimates, for instance, might offer an initial demonstration.

It is appropriate to view small-area data in a broad perspective that includes decennial census information, the geographic systems for collecting and manipulating information, and the administrative data that can provide more timely intercensal estimates. We begin, in the next section, by describing the geographic tools required for small-area data.

#### ROLE OF TIGER

The central repository of the Bureau of the Census's geographic information is the TIGER system, a geographic database, built from the U.S. Geological Survey's files of map sheets and the Bureau of the Census's file of feature names and address ranges. Building on an earlier large investment by the two agencies, the Bureau of the Census made a major commitment in the 1980s to updating and merging these two databases for use in the 1990 census. TIGER was first used by the Bureau of the Census in its field work in 1988 and 1989 and in local governments' work with the 1990 Precensus Map sheets. The revised TIGER data were then used for all the geographic work for the 1990 census.

But the TIGER database and product system did not end with the 1990 census. The geographic database needs to be kept current with the addition of new streets and recent address changes. Address ranges for rural areas also need to be added to the 1990 database, and new housing construction must be added as well. A database with up-to-date address information is essential for proper linkage to administrative records.

The cartographic quality of TIGER is important, as it is for any database. Unless there is precise definition of such geographic features as roads, rivers, and other physical structures, it is difficult to locate rural households. Physical structures with exact positions are also important for planning bridges, roads, postal routes, and other important infrastructure features. Good cartographic quality is also important for maintaining the accuracy of geographic overlays. If additional geographic features, such as telephone or electrical lines, are to be mapped correctly, they require high-quality geographic information in the original database. If it develops and maintains high-quality characteristics, TIGER could prove to be a powerful tool for geographic information.

The present TIGER program includes a system of exceptional value for the geographic coding and presentation of census data. An enhanced TIGER system could provide a framework for data collection down to the smallest geographic area. An up-to-date TIGER system would also be a significant administrative record system in itself and would support the data collection of census, survey, and administrative records. A maintained, enhanced TIGER system could, in fact, provide the base map for the nation's geographic statistics, including pertinent data on the movement of goods, services, and information products. An enhanced TIGER could also offer three benefits for decennial census operations: as a tool for taking a cheaper, more accurate census; as a building block for linkage to administrative record systems; and as a more effective basis for census-related sample surveys.

An enhanced TIGER is fundamental to a new small-area data program. If TIGER is to be used as the geographic database for the 2000 census, additions and corrections need to be made to TIGER within the next few years. However, enhancements to TIGER are also important to make use of TIGER, along with a national address file, with administrative records. The possibility of a new source of small-area data presumes that it will take many years to redesign and work with administrative records. During this time, TIGER and its associated address file could also be improved and enhanced.

#### ENVISIONING A NEW SMALL-AREA DATA SYSTEM

If the nation had an up-to-date TIGER and linked address file, the system would provide a continuous housing inventory. Such an inventory would permit states, towns, and local governments to generate small-area housing counts, on a very frequent basis. A national housing inventory, with linked small-area population data, would provide current information on thousands of small areas. It would link population changes in the intercensal period for areas in ways that are not currently available today, and it would provide more frequent estimates. If the geographic and address files are linked to attribute records (databases with such specific information as assessed value of the housing structure), other estimates could be produced. Linked housing characteristics, for example, could produce many current housing statistics for small areas on a timely basis.

A considerable amount of administrative data can provide small-area statistics without specific individual record linkage. Records for the Aid to Families with Dependent Children (AFDC) program usually have address information that can be readily geocoded, and the data can then be aggregated for small areas for analysis purposes. Such small-area data could facilitate policy planning, improve the targeting of public programs, make surveys more efficient, and improve the traditional census with more timely information.

There is an important distinction pertaining to the issue of linking administrative records. In one implementation, administrative records could be linked to residential addresses, for some geographic areas, in order to recreate existing census-type data. But, it is already common to geocode administrative records in order to provide small-area data. It should be recognized that both types of geographic linkage are possible, with merits and limitations for each.

We note again that decennial census information is, by definition, available only every 10 years. It is the nature, therefore, of decennial estimates to be several years (and sometimes more than 10 years) out of date when they are used. The chief advantage of using continuously collected administrative records is timeliness. Small-area statistics from available records could provide frequent, punctual estimates for policy and program planning.

As noted above, there are four concerns about developing a new basis for small-area data.

First, an operationally enhanced TIGER system with linked administrative records will pose privacy concerns. Although such a system would be maintained with precautions for the security of individual information and no individual information would be released, there will be worries about the maintenance of such a system by the federal government. There is need for further analysis of the privacy and confidentiality implications and security requirements of a system that uses administrative records. Second, administrative records do not contain the same or all of the information currently collected on and provided by the decennial census. Third, many administrative records systems are not comparable among various agencies, states, localities, or other divisions. Fourth, it is impossible to provide cost estimates at the moment. There would be costs for redesigning and improving administrative records and for the processing and linkage of records with a geographic database. The benefits of such a system include more timely information, including potential information from program data that are currently lacking. What value should be placed on information that may improve public programs, on being able to allocate public funds based on more up-to-date small area estimates? These are important issues for further study, and a key factor will be field experience with administrative records in conjunction with an enhanced TIGER system.

#### A CANADIAN EXAMPLE

An illustration may be instructive of how administrative records can be used to improve information for small geographic areas. Since 1982, administrative tax records in Canada have been used to provide annual estimates for relatively small areas (Statistics Canada, no date). Under the Statistics Act in Canada, Statistics

Canada, the national statistical agency, has access to administrative records. In particular, the Tax Act contains a provision for the transmission of copies of tax records held by Revenue Canada, the agency responsible for administering the corporate and personal income tax system in Canada, to Statistics Canada.

There are three differences in the personal tax records between Canada and the United States. First, in Canada, only individuals submit tax returns; there are no joint returns, as in the United States. Second, the Canadian tax system provides benefits to persons and families with low incomes, called refundable tax credits.<sup>1</sup> These refunds can be obtained by persons who paid no income tax. There is therefore a motivation for low-income persons to file a tax return, even though they paid no income tax. Third, some transfer payments are taxable in Canada. Taxable transfer payments include unemployment insurance benefits, Old Age Security benefits, and Canada Pension Plan benefits. Individuals with taxable transfer payments must also file tax returns.<sup>2</sup> One result of these differences is that a very high proportion of Canadians file tax returns, including low-income individuals who, if they were U.S. residents, would probably be nonfilers.

Statistics Canada receives the individual personal file, then edits and performs some imputations at the beginning of a 6-month processing cycle. During several processing steps, families are formed from the individual records. The address, sex, marital status, and age of individuals are noted in the records. Finally, in the last processing step, nonfiling dependents are imputed. For the 1991 tax year, personal tax returns were filed in April 1992, and the small-area data will be released by Statistics Canada in summer 1993. The statistical data for small areas include two different items. One item is data on families, including individual tax filers (living alone) and tax-filer families (husbands-wives and equivalents, single-parent families, and nonfamily persons). The second item is demographic information for both individuals and families. The demographic information consists of age, sex, family type, family size, and income information (total income, median income, and sources of income). Annual data are presently provided for the nation, its provinces and territories, for Census Metropolitan Areas (comparable to statistical metropolitan areas in the United States), and for Census Divisions (comparable to U.S. counties). Data are also produced for about 23,000 postal areas and about 5,000 communities that range in size from 100 to several million persons. The Canadian postal code zone is comparable to the U.S. nine-digit ZIP code area and, hence, is approximately a U.S. census block; however, Canadian postal codes are combined into larger areas if there are relatively few households in a code area. These data on population counts—with associated data on age, sex, family type, and income—have proved to be invaluable for timely small-area information.

Statistics Canada reports that the cost of the 1991-92 fiscal year program was about C\$1,780,000, or about 7 cents per record, for both tax filers and imputed records. In 1991-92, about C\$490,000, or more than one-fourth of the program costs, was raised through sales of the small-area data on individuals and families.

#### KEY ISSUES

The thrust of the previous argument is that consideration of the small-area data requirements for the census also should take into account the provision of estimates for the intercensal period. We close this discussion by noting some critical questions for the expanded use of administrative records.

The availability of intercensal estimates for small areas, first and foremost, provides more timely information. However, such information also has an important secondary impact: it decreases the intense demands placed on the decennial census which, at the moment, is the sole source for much of the nation's small-area data. Frequent estimates for small-area data might mean, in some instances, that the information would not need to be collected as part of the census. Thus, the first issue that must be addressed is what small area data require more timely estimates.

Data from administrative records differ in coverage, content, and quality from that collected in the decennial census. Discussion of small-area estimates from administrative records, therefore, needs to consider the data quality requirements of small area information. Would estimates from samples or statistical modeling suffice, especially if they provided up-to-date statistics?

<sup>1</sup> The U.S. tax system includes an earned income tax credit, but only poorer individuals with earnings file for a refund. All Canadian families with low incomes are eligible for a refundable tax credit.

<sup>2</sup> Unemployment and social security benefits are also taxable in the United States; however, individuals and families below a minimum income need not file a tax return.

Administrative data are collected, by their very nature, for use in programs and for managerial decisions. The timing of data collection and processing may vary greatly between record systems and agencies. Discussion of the use of administrative records requires investigation of the readiness of records, the required redesigning of records to improve their quality and to expand their content, and the length of time it would take to make them available for statistical analysis and small-area estimates. What can be said about the timeliness requirements for small-area data? If decennial census data are not frequent enough, how often are estimates needed?

As discussed above, some administrative records already have postal codes or other geographic references on them. These types of records could be used immediately, with no need for linkage to a geographic database or to other records, to provide estimates. For example, a school district that has a postal code associated with each student can prepare a map of the number of students by postal area without any additional geographic information. Other records lack a geographic reference, and some records may have only a street address. These types of records would have to be linked to a geographic reference system, such as TIGER. Once linked to a geographic database, the administrative data can provide various small-area estimates. But what levels of small-area data are required? Would census block-level estimates be sufficient for all purposes, including the development of alternative geographic boundaries by users?

The confidentiality of administrative records is an important concern. Federal, state, and local agencies have obvious public and legal requirements for safeguarding the individual confidentiality of their records systems. However, the confidentiality of records varies. Federal agencies, such as the Bureau of the Census, have special restrictions that protect the individual confidentiality of all the data collected. In contrast, information on housing structures, including date of construction, location, and assessed value, are routinely available for public inspection. What confidentiality requirements should be placed on small-area data derived from administrative records? If administrative data are linked for small-area estimates, should the requirements differ from those placed on current decennial census information?

The provision of small-area estimates using administrative records involves different government agencies and, possibly, different levels of government. Some of the ideas described above consider the use of housing information that is available only at the local level. Other administrative data exist only in state records. What are the government roles for a small area statistical system? Although the federal government might take a leadership role, what would be the nature of a state, local, and federal partnership?

#### REFERENCES

*Alexander, C.*

1993—Untitled [notes on continuous measurement designs]. Draft paper. Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

*Anderson, M.*

1988—*The American Census: A Social History*. New Haven, Conn.: Yale University Press.

*Baker, G.E.*

1986—Whatever happened to the reapportionment revolution in the United States? Pp. 257-76 in Bernard Grofman and Arend Lijphart, eds., *Electoral Laws and Their Political Consequences*. New York: Agathon Press, Inc.

*Barnett, L. D.*

1993—Comments on Congressional Research Service Report for Congress: Legal Issues for Census 2000. Paper prepared for the Panel on Census Requirements in the Year 2000 and Beyond, Committee on National Statistics, National Research Council. School of Law, Widener University.

*Bean, F.D., B. Edmonston, and J.S. Passel, eds.*

1990—*Undocumented Migration to the United States: IRCA and the Experience of the 1980s*. Washington, D.C.: The Urban Institute Press.

*Bryant, B. E.*

1993—Decision of the Director of the Bureau of the Census on whether to use information from the 1990 Post-Enumeration Survey (PES) to adjust the base for the intercensal population estimates produced by the Bureau of the Census. *Federal Register* 58(1) 69-78.



*Bureau of the Census*

1976a—Language Minority, Illiteracy, and Voting Data Used in Making Determinations for the Voting Rights Act Amendments of 1975 (Public Law 94-73). Current Population Reports, Population Estimates and Projections, Series P-25, No. 627. Washington, D.C.: U.S. Department of Commerce.

1976b—U.S. Census of Population and Housing: 1970. Procedural History. PHC(R)-1. Washington, D.C.: U.S. Department of Commerce.

1982—1980 Census of Population and Housing, Users' Guide, Part A. Text. PHC80-R1-A. Washington, D.C.: U.S. Department of Commerce.

1983a—Census of Population and Housing, 1980: Public-Use Microdata Samples—Technical Documentation. Washington, D.C.: U.S. Department of Commerce.

1983b—Stakeholders' Conference on Public Law 94-171 Program. 1990 Planning Conference Series No. 3. Washington, D.C.: U.S. Department of Commerce.

1990—Modelling Census Costs: Managing Risks and Opportunities. 21st Century Staff. Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

1992—1990 Census of Population and Housing, Guide, Part A. Text. 1990 CPH-R-1A. Washington, D.C.: U.S. Department of Commerce.

*Citro, C.F., and M.L. Cohen, eds.*

1985—The Bicentennial Census: New Directions for Census Methodology in 1990. Panel on Decennial Census Methodology, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

*Citro, C.F., and G. Kalton, eds.*

1993—The Future of the Survey of Income and Program Participation. Panel to Evaluate the Survey of Income and Program Participation, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

*Davidson, C.*

1992—The Voting Rights Act: A brief history. Pp. 7-51 in Bernard Grofman and Chandler Davidson, eds., *Controversies in Minority Voting—The Voting Rights Act in Perspective*. Washington, D.C.: The Brookings Institution.

*Durbin, T.N., and L.P. Whitaker*

1991—Congressional and State Reapportionment and Redistricting: A Legal Analysis. Congressional Research Service Report for Congress, 91-292-A. Washington, D.C.: U.S. Government Printing Office.

*Ehrenhalt, A.*

1983—Reapportionment and redistricting. Pp. 44-71 in Thomas E. Mann and Norman J. Ornstein, eds., *The American Elections of 1982*. Washington, D.C.: American Enterprise Institute for Public Policy Research.

*Ericksen, E.P., L.F. Estrada, J.W. Tukey, and K.M. Wolter*

1991—Report on the 1990 Decennial Census and the Post-Enumeration Survey, Appendix A: The Census Process. Report submitted by members of the Special Advisory Panel to the Secretary, U.S. Department of Commerce, Washington, D.C.

*Fellegi, I.P.*

1981—Comments. Pp. 51-68 in Congressional Research Service, *Using Cumulated Rolling Samples to Integrate Census and Survey Operations of the Census Bureau*. Prepared for the Subcommittee on Census and Population, Committee on Post Office and Civil Service, U.S. House of Representatives. Washington, D.C.: U.S. Government Printing Office.

*Gobalet, J.G., and S. Lapkoff*

1993—Voting Rights Act Issues in Political Redistricting. Paper prepared for the Population Association of America annual meeting, Cincinnati, Ohio. Lapkoff & Gobalet Demographic Research, Inc., Oakland, California.

*Grofman, B., L. Handley, and R.G. Niemi*

1992—*Minority Representation and the Quest for Voting Equality*. Cambridge, Eng.: Cambridge University Press.

*Henry, C.*

1989—The impact of new technology and new census data on redistricting in the 1990s. Pp. 67-74 in William P. O'Hare, ed., *Redistricting in the 1990s: A Guide for Minority Groups*. Washington, D.C.: Population Reference Bureau, Inc.

*Herriot, R.A., D.V. Bateman, and W.F. McCarthy*

1988—The Decade Census Program: A New Approach for Meeting the Nation's Needs for Sub-National Data. September 27. Unpublished manuscript, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

1989—The decade census program—a new approach for meeting the nation's needs for sub-national data. In Proceedings of the Social Statistics Section. Alexandria, Va.: American Statistical Association.

*Horvitz, D.G.*

1986—Statement to the Subcommittee on Census and Population, U.S. House of Representatives, May 1. Research Triangle Institute, Research Triangle Park, N.C.

*Issacharoff, S.*

1993—Comments on Congressional Research Service Report for Congress: Legal Issues for Census 2000. Paper prepared for the Panel on Census Requirements in the Year 2000 and Beyond, Committee on National Statistics, National Research Council. School of Law, University of Texas, Austin.

*Harlan, P.S., and P. McCrary*

1988—Book review: Without fear and without research: Abigail Thernstrom on the Voting Rights Act. *The Journal of Law & Politics* 4(4) 751-777.

*Kish, L.*

1981—Population counts from cumulated samples. Pp. 5-50 in Congressional Research Service, Using Cumulated Rolling Samples to Integrate Census and Survey Operations of the Census Bureau. Prepared for the Subcommittee on Census and Population, Committee on Post Office and Civil Service, House of Representatives. Washington, D.C.: U.S. Government Printing Office.

1990—Rolling samples and censuses. *Survey Methodology* 16(1) 63-79.

*Kominski, R.*

1985—Final Report—Documentation of Voting Rights Act Determinations. Unpublished memorandum to Paul Siegel. Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

1992—1992 Voting Rights Act Bilingual Ballots Determinations. Unpublished memorandum for the record. Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

*Kousser, J.M.*

1992—The Voting Rights Act and the two Reconstructions. Pp. 135-176 in Bernard Grofman and Chandler Davidson, eds., *Controversies in Minority Voting—The Voting Rights Act in Perspective*. Washington, D.C.: The Brookings Institution.

*Laney, G.P.*

1992—The Voting Rights Act of 1965, As Amended: Its History and Current Issues. Congressional Research Service Report for Congress, 92-578-GOV. Washington, D.C.: U.S. Government Printing Office.

*Lee, M.M.*

1993—Legal Issues for Census 2000. Congressional Research Service Report 93-177-A. Washington, D.C.: U.S. Government Printing Office.

*National Conference of State Legislatures*

1992—Customer Feedback for the 1990 Census Redistricting Data Program—A Preliminary Summary. Prepared for Discussion at the National Conference of State Legislatures Annual Meeting, Cincinnati, Ohio.

*NCSL Reapportionment Task Force*

1989—Reapportionment Law: The 1990s. Denver, Colo.: National Conference of State Legislatures.

*O'Rourke, T.G.*

1980—The Impact of Reapportionment. New Brunswick, N.J.: Transaction Books. Panel to Evaluate Alternative Census Methods

1992—Letter to B.E. Bryant, director, Bureau of the Census (December 14). Committee on National Statistics, National Research Council, Washington, D.C.

*Parker, F.R.*

1989—Changing standards in voting rights law. Pp. 55–66 in William P. O'Hare, ed., *Redistricting in the 1990s: A Guide for Minority Groups*. Washington, D.C.: Population Reference Bureau, Inc.

*Redfern, P.*

1989—Population registers: some administrative and statistical pros and cons. *Journal of the Royal Statistical Society A* 152 (Part 1) 1–41.

*Romig, C.L.*

1988—Evaluation of the 1980 Census and the Legislative Reapportionment Process. Denver, Colo.: National Conference of State Legislatures.

*Sailer, P., B. Windheim, and E. Yau*

1992—Toward a New Census Paradigm: Comparing Coverage of the Population in the Census and in Tax Documents. Draft paper and handout prepared for the American Statistical Association Annual Meeting, Boston, Mass. Statistics of Income Division, Internal Revenue Service, U.S. Department of the Treasury, Washington, D.C.

*Shapiro, G., and D. Kostanich*

1988—High response error and poor coverage are severely hurting the value of household survey data. Pp. 443–448 in *Proceedings of the Section on Survey Research Methods*. Alexandria, Va.: American Statistical Association.

*Statistics Canada*

Statistical Data from Canadian Personal Income Tax Records. Statistics Canada, Ottawa, Canada.

*Thernstrom, A.M.*

1979—The odd evolution of the Voting Rights Act. *The Public Interest* 55(Spring): 49–76.

1987—Whose Votes Count? Affirmative Action and Minority Voting Rights. Cambridge, Mass.: Harvard University Press.

*U.S. General Accounting Office*

1992—Decennial Census: 1990 Results Show Need for Fundamental Reform. GAO/GGD-92-94. Washington, D.C.: U.S. Government Printing Office.

#### BIOGRAPHICAL SKETCHES OF PANEL MEMBERS AND STAFF

CHARLES L. SCHULTZE is currently a senior fellow in economics at The Brookings Institution. Formerly, he was director of the U.S. Office of Management and Budget. His work has been primarily in macroeconomics and budgetary policy. He has often testified before Congress and in other forums on the statistical organization of the U.S. government. He served as chair of the Council of Economic Advisers in the Carter administration.

MARGO ANDERSON is a professor of history at the University of Wisconsin-Milwaukee. She has done extensive research in the area of census history and on the history of statistics. She has published widely on policy and historical development of the census, including *The American Census: A Social History* (1988).

CONSTANCE F. CITRO is a member of the staff of the Committee on National Statistics. She is a former vice president and deputy director of Mathematical Policy Research, Inc., and was an American Statistical Association/National Science Foundation research fellow at the Bureau of the Census. For the Committee on National Statistics, she has served or is currently serving as study director for numerous studies, including the Panel on Poverty and Family Assistance, the Panel to Evaluate the Survey of Income and Program Participation, and the Panel on Decennial Census Methodology. She is a fellow of the American Statistical Association.

DOUGLAS M. DUNN is a vice president with AT&T with responsibility for visual/multimedia communications strategy. After training in statistics, he worked at Bell Laboratories for several years before moving to AT&T, where he has served in many departments over the past 15 years. His major interest is data analysis, particularly time series and uses of data in business. He serves on many community boards, including United Way, Boy Scouts, Rotary Club, Fisk University, and the United Negro College Fund.

BARRY EDMONSTON is a study director with the Committee on National Statistics. He has been involved in demographic research and teaching at Stanford University and Cornell University, and he was senior research associate with the Program for Research on Immigration Policy at the Urban Institute. He has done re-

search on demographic methods, especially on the methodology of population projections, and on questions of immigration and immigration policy.

IVAN P. FELLEGI is chief statistician of Canada and deputy minister in charge of Statistics Canada. He has served as the president of the Statistical Society of Canada and the International Statistical Institute. Previously, he served on the Panel on Decennial Census Methodology of the Committee on National Statistics. He is a fellow of the American Association for the Advancement of Science and an honorary fellow of the Royal Statistical Society.

STEPHEN E. FIENBERG is Vice President for Academic Affairs at York University in Toronto and a statistician with long-standing interests in census activities and related methodology and data. As former chair of the Committee on National Statistics, he was instrumental in the creation of its Panel on Decennial Census Methodology in the 1980s. He has testified a number of times before the Subcommittee on Census and Population of the U.S. House of Representatives, most recently in March 1991, and he has written extensively on the 1990 census process.

CHARLES P. KINDLEBERGER is director of research for the St. Louis Community Development Agency. He is a user of census information and has a long-standing affiliation with professional associations, such as the Urban and Regional Information Systems Association and the American Planning Association. He is now serving as chair of the Information Technology Division of the American Planning Association.

MICHEL A. LETTRE is Assistant Director for Planning Data Services with the Maryland Office of Planning. He has served as the governor's chief staff person on the redistricting advisory committee and also coordinated the governor's 1990 census promotion campaign. His office serves as the principal agency for the distribution of census data within the state of Maryland under the Census Bureau's state data center program.

JAMES N. MORGAN is professor emeritus and emeritus senior research scientist at the University of Michigan, Institute for Social Research. He has been a professor of economics at the University of Michigan and is a pioneer in the use of survey research for understanding economic behavior. He is a member of the National Academy of Sciences.

WILLIAM A. MORRILL is president of Mathtech, Inc., a small, applied research organization in Princeton, New Jersey. Formerly, he was an official in the U.S. Department of Health, Education, and Welfare and the U.S. Office of Management and Budget. He serves on the Commission on Behavioral and Social Sciences and Education of the National Research Council and previously served on its Committee on National Statistics.

RICHARD F. MUTH is Fuller E. Calloway professor of economics at Emory University. His interests include spatial economics, the economics of housing and mortgage markets, and the economics of poverty and homelessness. He was formerly on the faculties of the University of Chicago, Washington University, and Stanford University. He has published extensively on spatial aspects of urban population distribution and housing markets.

JANET L. NORWOOD is a senior fellow at the Urban Institute. She is currently working on issues involving repairing and improving the federal statistical system, of which the Census Bureau is a large part. She spent more than 20 years in the federal statistical system, including 13 years as commissioner of the Bureau of Labor Statistics.

EROL R. RICKETTS is a senior research scholar at the City University of New York Graduate School's Center for Social Research. Formerly, he was a visiting scholar to the Russell Sage Foundation, and at the Rockefeller Foundation he administered policy analysis, research, and fellowship programs on urban poverty and the underclass. He has testified before Congress on race, ethnicity, and ancestry questions in the 1990 census and before the New York State Advisory Committee to the U.S. Commission on Civil Rights regarding the undercount in the 1990 census.

TERESA A. SULLIVAN is a professor of sociology and law and the Associate Dean of Graduate Studies at the University of Texas at Austin. She serves on the Bureau of the Census's advisory committee on population statistics, which she chaired in 1991-1992. She has previously served on National Research Council panels on immigration statistics and on technology and women's employment.

KARL TAEUBER is a social demographer at the University of Wisconsin-Madison, where he is affiliated with the Center for Demography and Ecology and the Institute for Research on Poverty. He teaches courses on race relations and population problems and does research on trends in school and residential segregation.

JAMES TRUSSELL is professor of economics and public affairs at Princeton University, director of the University's Office of Population Research, and associate



dean of the Woodrow Wilson School of Public and International Affairs. His publications are primarily in the areas of demographic methodology and reproductive health. He served for 6 years on the Bureau of the Census's advisory committee on population statistics, and he has served on many National Research Council panels.

Mr. SAWYER. Thank you very much. That was a marvelous journey through a much longer and complex document. It is encouraging to think that it is really only the interim report on work that continues. But clearly you have touched on an awful lot of issues that are of concern to this panel and to other Members across the Congress.

You mentioned in the appendix on small area data needs, the importance of the census as a benchmark in using administrative records. Do you have any idea how much data we would need to retain to serve as an effective benchmark? Is that the sort of thing you will be looking at?

Mr. SCHULTZE. Well, let me say a few things and I might want to ask Barry to elaborate, if he would. I think you have to make several distinctions. One, it depends on whether you are talking about the use of administrative records for intercensal estimates in a system where the census is basically unchanged, or whether you are talking about the use of administrative records as a substitute.

Mr. SAWYER. No, I am talking about a system in which the census remains the cornerstone. What we are trying to do is to come to some sense of how much of a cornerstone we need.

Mr. SCHULTZE. Well, let me make a stab at two things certainly. One—

Mr. SAWYER. I am not even asking you the answer to the question, but whether or not we are—

Mr. SCHULTZE. No, I understand. I am trying to look at the areas you have to look at and I can think of two of them out of maybe 20. One is an address list. Some sort of geographical master housing list, address lists or something like that I think is necessary.

Now, it can be in various forms, maybe just ranges of addresses. But in any event, it has got to be there in order to assign administrative records to a geographical area. Now, there for some kind of needs as long as you know the city you get the data from, maybe you don't care about smaller detail.

I am putting that aside. I am talking about the kind of small area data where it is important to know at some significant level of geographical detail where the person or family or unit is located. If so, you have got to have an address list and you have got to match them. If you want to be able to cross-tabulate, you have got to match the data address by address. That is, if you want to be able to take information about income and match it to information about housing, you have got to do it that way.

Now you can begin to substitute certain kinds of small area averages, you can get a block average.

Mr. SAWYER. Let me ask you, when you are talking about income, is income data perhaps so flawed that it cannot serve as an effective benchmark from within the census and that we are better off looking to administrative records, estimates built off of those on a more continuous basis throughout the decade?

Mr. SCHULTZE. Well, my limited experience in working with income data suggests that you are in serious trouble if you rely

only, say, on tax records, but you are also in serious trouble if you rely only on census records. And the best information I know, it is also controversial because you have to make some judgments, is where you merge the two sets of data.

Now, we do that now in the aggregate. We have various estimates of family income. But it turns out that ones that use only Census or only IRS are going to be seriously flawed. So I would say you have got to mix them.

And if you didn't have census-like information—I didn't say census information, but census-like information that might, for example, come out of an annual survey rather than the census itself, but census-like information—I would be worried that the administrative records would give you flawed results.

On the other hand, I must confess that using only census information for income data also gives you flawed results because the responses tend to be much less, we hope, careful than when you are sitting there facing the tax collector. So there is no easy answer to that.

But my point is, at a minimum, you have got to do that matching of addresses. Then secondly, the question is: What do you need by way of a sampling frame, what do you—how much do you need that decennial census income and other data as a framework within which to do, for example, sampling. And here I know it is important, but I don't quite know what the trade-offs are yet.

Mr. SAWYER. That was the point of the question I asked you just prior to this, how much of a base do you need in that cornerstone.

Mr. SCHULTZE. The experts tell us and the numbers vary, I have two surveys in mind, one, the consumer expenditure survey, which is an ongoing operation that is used initially for the consumer price index, but actually used for all sorts of other information.

And we are given to understand it would cost about 30 percent more to do that if you didn't have the census sampling frame because they would have to have a much bigger survey to know that they were really catching it right when you have got to break it down.

On the other hand, again, if I remember correctly, to do the CPS, the Current Population Survey—the one that comes out each month and that then is done annually in March, this is about a 5 percent additional cost if you don't have a census frame for sampling. In the aggregate, when you add all these up, I don't know, it is a substantial difference if you don't have the Census, mainly in terms of doing a much bigger survey each time.

And then you have got to ask yourself, well, if you are going to do the census, even if it costs you in terms of response, maybe it is still worth it to a point, if you get information that saves you costs another way. And I have to confess, while in general we know all this and we recognize the trade-offs, we haven't yet done the work and I am not sure we are going to be able to do it fully.

Mr. SAWYER. I am not sure that is even a reasonable expectation of the work of the panel, but to define the problem and shape the direction the answers must take is really the important thrust of the work.

Mr. SCHULTZE. A very important part of all of this is that once you are going to do a census, if you say you have got to do it for

various purposes, then how much additional cost is it to hang something else on it? And my impression is that really, if you look at it item by item, it isn't very expensive, but what you have to be careful about is making it too complicated. That is not a mathematical answer.

Mr. SAWYER. I think that is right. I appreciate your caution in drawing too close a linkage between content and accuracy. And yet the Bureau's own measurements really do demonstrate that reducing content, for example, on the short form, really does improve mail response rates.

Mr. SCHULTZE. The only thing I would add, just to make sure we pull all of the pieces in, that interestingly enough, it might be worthwhile reducing response rates if the cost of foregoing that data, in terms of what you have got to do elsewhere, is very large. So it is an even more complicated trade-off.

Mr. SAWYER. You also suggested something else that I think is worth looking into with some care. And I don't know how to measure it, but I am hopeful that your folks can give us some guidance. And that is that holding content steady may nonetheless show great variations in response rates, which is probably more a measure either of the complexity of the format or a diminishment in public tolerance.

And I don't know how you measure that, but it seems to me that that is a very real function that we need to take into account. I just hope that you will explore that further in the next 18 months, because I think it is an important question.

The other thing is that I completely understand what you are talking about in terms of the marginal differences in reducing content. Savings would be minimal, at least direct savings, but there are indirect costs certainly in terms of pursuing nonresponse follow-up in traditional ways, which just escalate costs enormously and very rapidly.

It is obviously the point that you are getting to with regard to sampling follow-up for response. I hope that we can pursue that in even greater detail.

Mr. SCHULTZE. Yes, sir.

Mr. SAWYER. With that, let just let me say thank you. We will look forward to hearing from you again sooner than probably any of us might anticipate.

For the record, would you identify your colleague?

Mr. SCHULTZE. Excuse me, I should have. This is Dr. Barry Edmonston who is, of course, director of our staff, of the study. I apologize.

Mr. SAWYER. It is not so much for me as our friend at the front table.

Mr. SCHULTZE. I am so used to having him next to me that it is natural.

Mr. SAWYER. Thank you very much. I appreciate your being here.

Our second witness today is Dr. Harry A. Scarr who is Acting Director of the Bureau of the Census. I think perhaps the only job I can think of over there that would be harder than to be the new director designate is to continuously sit there as the acting director. I have great admiration for you.

I understand that your daughter, Rachel, is with you.

Mr. SCARR. Yes, she is, Mr. Chairman.

Mr. SAWYER. I am pleased to welcome her here. As usual, the entire text of your remarks will be included as part of the record and you are free to proceed as you will.

**STATEMENT OF HARRY A. SCARR, ACTING DIRECTOR, BUREAU OF THE CENSUS; ACCOMPANIED BY SUSAN MISKURA, CHIEF, 2000 CENSUS RESEARCH AND DEVELOPMENT STAFF**

Mr. SCARR. Thank you, Mr. Chairman. Seated with me is Susan Miskura, who is the chief of the 2000 Census Research and Development Staff at the Bureau.

It is a pleasure to be here today to testify before the subcommittee on the status of Census Bureau planning for the 2000 census. Since I last testified on this subject on March 2, the Bureau has completed criteria for assessing design alternatives, accelerated the assessment process, and issued design alternative recommendations ahead of schedule.

This acceleration means that we and our stakeholders will be able to have more time to review the recommendations about which design features should be part of the 1995 test. It also means that we will be able to focus sooner and more clearly on planning for this important test.

Mr. Chairman, with completion of the design alternative recommendations, we have reached a major milestone in moving toward our goal of designing a census for 2000 that is responsive to changing times.

When we originally presented our research program to this subcommittee and to its staff, then Associate Director Robert Groves, remarked that it was very likely that the design for the 2000 census would not resemble any of the 14 designs then being considered. The 14 designs were meant to be vehicles to organize our research in a systematic fashion.

Now we are moving beyond those 14 designs, and for the 1995 test we will use the best features and options from the designs as building blocks as we proceed to assemble the best census design.

From the tests of features in 1995 and from the inputs to this very open process we have maintained, we will determine the design for the 2000 census. Our recommended options for the 1995 test mean that we will be developing fundamentally new options for the design of the census.

To come full circle, Mr. Chairman, from the features and options we recommend for testing in 1995, it will be possible to construct a census design that is like, say, our matrix sampling design, or like our high tech census, or like our voting-rights-data-only census, or like our continuous measurement census—like, but not identical to any of these.

Let me now turn to the issues you asked me to address. First, criteria for assessing design alternatives.

On March 25, 1993, we published a notice in the Federal Register describing criteria, 5 mandatory and 13 desirable, for assessing design alternatives and asking for comment. As a result of our review of responses to this Federal Register notice, including one



from you, Mr. Chairman, we reduced the number of desirable criteria from 13 to 11, moved one criterion from the desirable to the mandatory category, and made minor changes to three other criteria.

The desirable criterion that we made mandatory states that a design should be assessed for its promise to reduce the differential undercount. This change reflects the many comments that we received recommending that this criterion be made mandatory. These criteria complement the three overarching goals which continue to drive our design efforts: (1) to reduce differential undercount, (2) to contain costs, and (3) to continue to keep the process open.

Next, the status of 1995 test preparations. The 2000 Census research and development staff completed and released recommendations on each of the 14 alternative designs, as I indicated, ahead of our planned schedule. By issuing recommendations now, we allow ample opportunity for a wide public review this summer before locking in by this September final specifications for the 1995 test.

Let me provide a summary of the options we plan to test in 1995, according to our current plans. First, we recommend developing and using a full range of primary options for responding to the census in the 1995 test.

Second, we recommend continued research on the attributes of a census with no follow-up or with a shortened follow-up period and the use of multiple data collection modes to complete the follow-up.

Third, we are recommending testing and evaluating the use of sampling for nonresponse follow-up in the 1995 test.

Fourth, we recommend the use of the multiple sample forms in the 1995 test. Also called matrix sampling, this methodology differs fundamentally from the approach of recent censuses which used only one sample questionnaire.

Fifth, we recommend incorporating integrated coverage measurement into the 1995 test. Specifically, the 1995 test will be a one-number census test producing a single set of census results.

Six, certain features merit testing in 1995 because of both their role in any census design, as well as their potential for fundamental changes. Those features will be tested.

Seven, we are recommending the use of the master address file, which is a housing unit list with administrative records, in the 1995 test.

Eighth, the 1995 test will enable us to develop accurate and cost-effective methods for the decennial year portion of a continuous measurement system.

Now, I will discuss the status of census content development.

In December 1992, the Office of Management and Budget asked 28 Federal departments and agencies to describe in detail their perceived requirements for topics to be included in the 2000 Census. We are summarizing and evaluating the responses we have received to date.

Once a decision is made on the scope of content, the next phase of the development program must focus on identifying specific topics and questions that are appropriate within that scope. That phase will involve both Federal and non-Federal data users and will take several years. By law, we must report to the Congress on

proposed topics no later than April 1, 1997, and on proposed questions no later than April 1, 1998.

Next, I will turn to the status of research activities. In my testimony on March 2, I summarized research activities in four areas: (1) barriers to enumeration, (2) administrative records, (3) improving mail response, and (4) technology. Let me briefly provide you with progress in these same four areas and then describe in more detail additional areas of research such as sampling.

**Barriers to enumeration.** Since March, work has progressed in this area on the development of specific tools to be included in a tool kit of special enumeration methods. We have begun work to test our ability to mail out Spanish language questionnaires to areas with a high concentration of Hispanic respondents.

We are beginning work to gain a better understanding of living arrangements of Asians, Pacific Islanders, American Indians, and Alaska natives. We are reviewing the recommendations we heard at the Richmond undercount conference to determine what additional research should be added to the program.

**Administrative records.** Administrative records research continues with a major focus on identifying uses of those records for the 1995 test. Some potential uses under investigation include improving the coverage of the master address file, updating mailing lists for special places such as colleges and hospitals, using administrative records to collect information about persons living at special places, and exploring the integration of sampling for nonresponse follow-up with the use of administrative records as a source of information for persons not easily enumerated.

In July, the Bureau will hold an interagency conference on access and other barriers to the statistical use of administrative records. Work also continues on matching administrative records files such as voter registration, school district and tax assessment files, to data collected in the special censuses of Godfrey, IL, and South Tucson, AZ.

**Improving mail response rates.** We are nearing completion of the data collection phase of the mail and telephone mode test. Our preliminary results show that offering households the option of responding by telephone during subsequent mail contacts, after a respondent has had an opportunity to examine the census form received in the mail and has not yet completed it or returned it by mail, does improve response rate slightly.

We will combine these findings with other response research results to maximize response to the census. This July, we will mail out questionnaires for our appeals and long form experiment.

**Technology.** Census Bureau technology research is looking at ways of providing householders multiple means of responding to the census. Recently, outside experts under contract to the Bureau provided us with a report assessing various household based technologies and their availability and viability as data collection modes in 2000. We are in the process of evaluating that report.

At the March 2 hearing, I noted the review of our system for putting data on the census questionnaires into computer readable format for speedy processing. We have been looking at promising new commercially available technology that has the ability to scan

properly designed questionnaires and electronically interpret many answers.

**Sampling research.** We are conducting extensive research on the use of sampling in the 2000 census in four areas: Coverage measurement, truncated census, sampling for nonresponse follow-up, and matrix sampling. We are conducting research to assess the meaning of certain key concepts that respondents apply in determining household membership to assist us in our coverage research.

**Continuous measurement.** We have begun developing and evaluating a prototype design for a continuous measurement census system.

**Status of the Bureau and Postal Service joint cooperative work.** Staff from our two agencies have met several times, most recently on March 26, 1993. We have agreed to conduct a test study in which the U.S. Postal Service will perform an automated match of selected addresses from our 1990 census address control file against address information they have compiled.

The staff of this subcommittee convened a meeting of U.S. Postal Service and Census Bureau staff on April 19 to discuss progress on the cooperative program. We were encouraged by the meeting and hope the subcommittee staff was reassured that the two agencies should be able to work together to develop a strategy for updating our address list and the TIGER database.

The interagency Joint Committee for Census Planning which was established by the Bureau and the U.S. Postal Service in 1990, met on March 3 to discuss sharing of address information, identification of vacant units and rural addressing issues. The joint committee has established two subcommittees to explore specific areas of cooperation. One is looking at encouraging local efforts to convert from rural style to city style addresses. The second is looking into possible changes in legislation that would allow the U.S. Postal Service and/or the Census Bureau to share address information.

Mr. Chairman, as you can see, we are conducting an extensive research and development program for the 2000 census. That concludes my testimony. I will be happy to answer any questions.

[The prepared statement of Mr. Scarr follows:]

PREPARED STATEMENT OF HARRY A. SCARR, ACTING DIRECTOR, BUREAU OF THE  
CENSUS

#### INTRODUCTION

Mr. Chairman, it is a pleasure to be here today to testify before this Subcommittee on the status of Census Bureau planning for the 2000 census. I am pleased to report that since I last testified on this subject on March 2, the Census Bureau has made significant progress in 2000 census planning. We have completed the criteria for assessing design alternatives, and accelerated the assessment process, and our 2000 Census Research and Development staff issued their design alternative recommendations ahead of schedule. We originally intended to issue them on a flow basis between March and this September, but we issued all 14 by May 17.

This acceleration means that we and our stakeholders will be able to have more time to review these recommendations about which design features should be part of the 1995 test. It also means that we will be able to focus sooner and more clearly on planning for this important test.

Mr. Chairman, with completion of the design alternative recommendations, we have reached a major milestone in moving toward our goal of designing a census for 2000 that is responsive to changing times. When we originally presented our re-

search program to this Subcommittee and to its staff, then Associate Director Robert Groves remarked that it was very likely that the design for the 2000 census would not resemble any of the 14 designs being considered. That was because the designs were not to be prisons, but were to be heuristic devices to move the design process forward.

The 14 designs were meant to be vehicles to organize our research in a systematic fashion. They served that purpose well. Now, we are moving beyond those 14 designs and, for the 1995 test, will use the best features and options from the designs as building blocks as we proceed to assemble the best census design.

To come full circle, Mr. Chairman, from the features and options we recommend for testing in 1995, it will be possible to construct a census design that is like, say, our matrix sampling design, or like, say, our high tech census, or like, say, our voting-rights-data only census, or like, say, our continuous measurement census. Like but not identical to any of these. For as I noted a moment ago, the designs have always been heuristic devices to assist understanding of the essentials of census taking, and not straightjackets to impede such efforts. From the tests of features in 1995 and from the inputs to this very open process we have maintained, we will determine the design for the 2000 census.

Our recommended options for the 1995 test mean that we will be developing fundamentally new options for the design of the census. Even with a continuous measurement census, which I will discuss in more detail later, there would likely be a need for a census that collects minimal data in the decennial census year. Many of the options we recommend testing in 1995—like the Master Address File, multiple response modes, matrix sampling techniques, and the use of administrative records—would also apply to a range of intercensal activities that could be part of a continuous measurement system.

Let me describe how we are continuing to keep this an open process, and then I will address the specific issues you raised in your letter of invitation.

The three committees of the Task Force for Designing the Year 2000 Census and Census Related Activities for 2000-2009 continue to meet regularly. The Technical Committee has been meeting twice monthly to deliberate the various alternative designs as well as other research that is important for all designs. The Policy Committee, which has met twice since the last hearing, is working on Federal data needs and content for the 2000 census. The Advisory Committee met in March and is scheduled to meet for 2 days in June. At the June meeting, that Committee will take up the recommendations on the 14 design alternatives.

We are meeting regularly with members of the two National Academy of Science panels. We have been providing information to, and consulting with, the four working groups of the Panel to Evaluate Census Methods. These working groups are investigating: (1) a head count in the census year with other data collected over the decade, (2) alternative ways of responding to the census, (3) administrative records and lists, and (4) sampling and statistical estimation for coverage and differential coverage improvement.

We attended a meeting in April sponsored by this Subcommittee to discuss the status of activities involving the Census Bureau and the U.S. Postal Service related to TIGER File and address list updating, and on May 12 we briefed subcommittee staff on plans for the Appeals and Long Form Experiment that we will conduct this Fall. I will discuss this test and the status of our activities with the U.S. Postal Service later.

Early this month, the Census Bureau held a "Research Conference on Undercounted Ethnic Populations" in Richmond, Virginia to gain insight into, and hear proposals for, reducing the undercount of various populations. The purpose of the conference was to better understand the extent and causes of the undercount in 1990 and to identify ways to reduce it in 2000. Participants included academicians, members of our minority advisory committees, representatives from the public and private sectors, and Census Bureau personnel.

Mr. Chairman, we appreciate the support you lent to this endeavor by being the conference dinner speaker and by the participation of your staff at the sessions. We are now reviewing ideas from this conference for additional areas of research that should be incorporated into our 2000 census research and development program.

Let me now turn to the issues you asked me to address. I shall first describe the development and use of criteria for assessing the design alternatives, then turn in more detail to the design alternative recommendations and the status of preparations for the 1995 field test, then provide an update on census content development, followed by a discussion of the status of research in various areas. Finally, I will provide an update on joint activities between the Census Bureau and the U.S. Postal Service.



# CRITERIA FOR ASSESSING DESIGN ALTERNATIVES

To assist our efforts in assessing the design alternatives and selecting the design options to test in 1995, we developed, in consultation with stakeholders, the Office of Management and Budget, and the Department of Commerce, a set of proposed criteria. On March 25, 1993, we published a notice in the Federal Register describing these criteria—5 mandatory and 13 desirable—and asking for comments by April 24.

We received and reviewed over 140 responses to the Federal Register notice, including several received after the deadline. Many of the comments were valuable and directly relevant to helping us refine the criteria for assessing designs. After careful review of all responses, we have prepared a final version of the criteria.

Parenthetically, Mr. Chairman, we appreciate the very thoughtful comments you sent on the criteria and we have made every effort to consider your comments, and those of others who responded, as the final criteria were developed.

As a result of our review of all these responses to the Federal Register notice, we eliminated two of the desirable criteria, moved one criterion from the desirable to the mandatory category, and made minor changes to 5 other criteria. We changed the title of one of the desirable criteria to be responsive to the large number of comments we received that there is a critical need for small area socio-economic data that are only available from the decennial census. The criterion now reads, "Provide small area data that the census is uniquely capable of supplying."

One desirable criterion we deleted specified that methods and procedures should be understandable and credible to the public. Through the open process we have for 2000 census design and planning, we believe that stakeholders will be able to understand the methods and procedures that we will use in 2000. The second criterion we deleted said that a Federal statistical structure should be in place to support the census. We deleted this criterion because it implied Census Bureau control over activities for which the Bureau does not have ultimate responsibility.

The desirable criterion that we made mandatory says a design should be assessed for its promise to reduce the differential undercount. This change reflects the many comments that we received recommending that this criterion be mandatory. Reducing the differential undercount for the 2000 census remains one of our highest priorities. Within the next month, we intend to publish the final version of the criteria in the Federal Register with a summary of responses and an explanation of any changes.

We have already used these final criteria to assess the 14 design alternatives and their component features. The design alternative recommendations document our assessment of the design alternatives against these criteria.

I want to emphasize that these criteria are guidelines for assessing designs and design features. As such, they complement—and in some instances reinforce—the three overarching goals which drive our design effort and which will drive our future operational planning effort. As I noted at the March hearing before this Subcommittee, these goals are:

- 1) To reduce differential undercounts—that is, the different rates at which population groups and geographic areas are covered by the census—and make the census more accurate overall,
- 2) To contain costs, and
- 3) To keep the process open.

## DESIGN ALTERNATIVE RECOMMENDATIONS AND STATUS OF 1995 TEST PREPARATIONS

The 2000 Census Research and Development Staff completed and released recommendations on each of the 14 alternative designs well ahead of the planned schedule. We accelerated our work to complete the recommendations because of the concern you expressed about the pace of our planning process, Mr. Chairman, and because of the urging of others to move quickly to isolate the workable components of the 14 alternative designs and develop a plan for testing them in 1995. By issuing recommendations now, we are allowing ample opportunity for a wide public review this summer before locking in, by this September, final specifications for the 1995 test. We are distributing the recommendations to our Task Force committees, to members of the National Academy of Science panels, to members of Congress and their staff, and to other key stakeholders, as well as to all others who express to us a wish to review them.

Each design alternative recommendation describes the range of options for the eight distinguishing features of taking a census and assesses that design in terms of the mandatory and desirable criteria. The recommendation describes why each of these options will or will not be tested in the 1995 test and cites the research or

other results that led to that choice. For options not selected, we recommend whether research should continue for further development beyond the 2000 census.

The design alternative recommendations provide our specific recommendations about each of the 14 full designs. But most importantly they document the combinations of options that we recommend be tested and evaluated in the 1995 test. The single design for the 1995 test will be a combination of the most promising options from the 14 design alternatives. We will "test" some of the options in a developmental sense; that is, not in a comparative, experimental way, but to gain information on cost, to measure accuracy, and to identify variations that are the most effective. For other options, we will build into the test some panels or simulations to determine the best variation of the option, or choice of options, to select. In some instances, we must still identify basic approaches to the options between now and September. The key concept, however, is that we will test in 1995 the most promising options from among the original 14 designs that may be combined into a number of "hybrid" designs from which to determine the 2000 census design.

These recommendations are subject to public review through the summer and will be amended as appropriate based on that review. Nevertheless, because we have kept the process open and are aware of many stakeholder concerns already, we have sufficient confidence in the major thrust of these recommendations to proceed both to begin basic operational planning for the 1995 test, and to continue research on those options where further definition is required.

It is well to remind ourselves once more, Mr. Chairman, that the critical goal of the 1995 test is to provide information to determine, by December 1995, the final design for the 2000 census. This goal will be met by determining how best to accomplish the eight distinguishing features of a census, as well as how best to implement the additional features that are important for any census design. Our test objectives will be to provide information on the cost, accuracy, and feasibility of the most promising ways to accomplish this.

We plan to select sites for the 1995 test by the end of October 1993. Our current plan is to undertake a large-scale test in three urban sites and one rural site. Having diversity in the sites will aid in generalizing test results to the national level. Let me provide a summary of the options we plan to test in 1995, organized by census feature:

#### *Primary Response Options*

We recommend developing and using a full range of primary options for responding to the census in the 1995 test. This is a considerable change from previous censuses, where the only option available to most of the population was to complete and return a questionnaire by mail. We will continue to use mail, with improvements as identified in the research program to date (specifically, multiple contacts with respondents, a more user-friendly questionnaire, and questionnaire simplification); but we will also use face-to-face enumeration and telephone interviewing as primary response modes where appropriate. Other electronic modes we can use in the test will be identified shortly. We plan to use assistance centers where people can obtain assistance in completing questionnaires, language aids, and multiple locations for obtaining blank questionnaires, as well.

#### *Nonresponse Follow-up*

This feature involves two ideas. The first is whether we conduct nonresponse follow-up at all. At this time, we recommend continued research on the attributes of a census with no followup or with a shortened follow-up period. We call this a "truncated" census. If research between now and September determines that this is a viable alternative, we will test it in the 1995 test.

The second idea is the use of multiple data collection modes to complete the follow-up. We currently recommend using the telephone, and using computer assisted interviewing.

Both of these approaches represent a fundamental change from past census processes, where we conducted a complete follow-up and used personal interviewing, with paper and pencil, as the basic methodology.

#### *Sampling for the Count*

We are recommending testing and evaluating the use of sampling for nonresponse follow-up in the 1995 test. We are currently looking into the design for, and the sample size required by, such a test. We will make a final decision on this test by September. We also recommend the use of sampling for coverage improvement purposes during the test. We will develop various scenarios for this option by September—options which can be incorporated into operational planning. Use of sampling

in these ways is a fundamental change from the approach used in recent censuses, when we attempted to count all persons and housing units.

It is important to note that sampling for nonresponse follow-up still involves the construction of a complete list of housing unit addresses, and an attempt to get mail returns from as many of those addresses as possible. The logic of sampling for nonresponse follow-up is to follow up only a sample of the addresses for which questionnaires were not returned, and to make inferences based on information collected from the sample for those addresses not in the sample and not visited during nonresponse follow-up. The published data on the number of persons and the characteristics of those persons from such a census would consist of a mixture of data mailed in by householders, data collected from addresses in the nonresponse follow-up sample, and data inferred for nonsampled addresses.

Legislation may be required to allow us to use sampling in arriving at census counts.

#### *Sampling for Content in the Decennial Year*

We recommend the use of "multiple sample forms" in the 1995 test. Also called "matrix sampling," this methodology differs fundamentally from the approach of recent censuses, which used only one sample questionnaire. We will test matrix sampling in the 1995 test to learn how to implement the procedures, and how to produce estimates from such a sample.

#### *Statistical Estimation*

We recommend incorporating integrated coverage measurement into the 1995 test. Specifically, the 1995 test will be a "one-number census" test producing a single set of census results. Research between now and September will determine the specific measurement methodology or methodologies to be used.

#### *Features Important for Any Design*

Certain features merit testing in 1995 because of their role in any census design, as well as their potential for fundamental change. We are now determining which such features to test in 1995.

#### *Use of Lists*

The Census Bureau, in cooperation with the U.S. Postal Service, is planning to create and continually update a Master Address File integrated with the TIGER geographic data base. This activity represents a change from our traditional approach to creation of a census address list, which has been to compile the address control list from scratch before each census. We are recommending the use of the Master Address File, which is a "housing unit list with administrative records," in the 1995 test. The administrative records will, at a minimum, be those of the U.S. Postal Service; files from the localities where the test is conducted; and available lists of telephone numbers.

#### *Data Collection Outside the Decennial Year*

We are committed to designing and testing a program that can produce data continuously throughout the decade between censuses. Collecting data with such a program would be a significant departure from collecting "long form" data from a sample as an integral part of the decennial census. The 1995 test will only enable us to develop accurate and cost-effective methods for the "decennial year" portion of a continuous measurement system. Because of its significance, I will discuss continuous measurement in more detail later.

#### *Content In The Decennial Year*

Content testing is not an objective of the 1995 test. 1990 census questions are adequate for purposes of the 1995 tests. However, because of the importance of content, and in response to your request, let me discuss the content development process for the 2000 census in some detail.

#### STATUS OF CENSUS CONTENT DEVELOPMENT

In December 1992, the Office of Management and Budget (OMB) asked 28 Federal departments and agencies to describe in detail their perceived requirements for topics to be included in the 2000 census. OMB's original target for receipt of agency input was mid-February. That date has been extended. OMB has received responses from most of the 13 Federal agency members of the Policy Committee of the Task Force and 4 responses from the 15 other agencies surveyed. OMB has shared those responses with the Bureau. We are summarizing and evaluating the responses received to date. We are paying particular attention to statutory citations that direct-

ly or indirectly require the use of decennial census data items, and to the level of geographic detail for which data are required. This process will continue as we seek to link these requests to data needs expressed by non-Federal data users.

However, we are recommending that the 1995 test census collect the same 100-percent and sample data as the 1990 census. The only modifications we will consider to the 1990 questions are those that result from our questionnaire simplification and coverage research. This allows us maximum flexibility in performing the 1995 test census. OMB, in working with the Policy Committee of the Task Force, hopes to complete its process by September 1993, and identify Executive agency data needs to determine the scope, or overall level, of content. As various stakeholders have pointed out, this may curtail adequate consideration of non-Executive data needs.

Based on the 1995 test results and other research conducted between now and the end of 1995, we will know much more about specific data needs, the extent to which and how they can be met through the use of a variety of census procedures, including a continuous measurement system, and many other factors that will affect the ultimate 2000 census design choice.

Once a decision is made on the scope of content, the next phase of the content development program must focus on identifying specific topics and questions that are appropriate within that scope. That phase will involve both Executive and non-Executive data users and will take several years. By law (Title 13, U.S. Code), we must report to Congress on proposed topics no later than April 1, 1997, and on proposed questions no later than April 1, 1998.

Let me once more emphasize as strongly as possible that the proposal to use the 1990 census content in the 1995 census test does not mean that this will be the scope of content for the 2000 census. Rather, it recognizes that decisions on content can and should be made in 1995 as part of the final choice of the 2000 census design, and as a result of a distinct, but parallel, coordinative process.

#### STATUS OF RESEARCH ACTIVITIES

In my testimony on March 2, I summarized research activities in four areas: barriers to enumeration, administrative records, improving mail response, and technology. Let me briefly provide you with progress in these same four areas and then describe in more detail additional areas of research, such as sampling.

##### *Barriers to Enumeration*

Since March, work has progressed on the development of specific tools to be included in a tool kit of special enumeration methods. The tool kit approach will give us more flexibility than in any previous census to meet special enumeration requirements of specific areas. Development of these tools began with a series of workshops involving Census Bureau headquarters and regional office staff who worked on the 1990 census. We held the first workshops in April and will continue them into this July.

We have begun work to test our ability to mail out Spanish language questionnaires to areas with a high concentration of Hispanic respondents. The main objective of this test, which we will conduct in October, is to determine if making Spanish-language questionnaires available by mail can improve the mail-response rate in largely Hispanic areas.

We are beginning work to gain a better understanding of living arrangements of Asians, Pacific Islanders, American Indians, and Alaska Natives. At the beginning of this month, we began data collection for our Living Situations Survey (LSS). This survey is based on a national, stratified sample of housing units with an oversampling of minority (Black, Hispanic, Asian) and rental housing units. We will collect descriptive data on new or changed household composition, mobility, residency patterns, and attachments of persons to households and other places that we must include in our enumeration and residence rules. We expect to complete data collection for this survey by mid-August.

We are also reviewing the recommendations we heard at the Richmond Undercount Conference to determine what additional research should be added to the program.

##### *Administrative Records*

Administrative records research continues with a major focus on identifying uses of administrative records for the 1995 test. Some potential uses under investigation are: (1) improving the coverage of the Master Address File, (2) updating mailing lists for special places, such as colleges and hospitals, (3) using administrative records to collect information about persons living at special places, and (4) exploring the inte-



gration of sampling for nonresponse follow-up with the use of administrative records as a source of information for persons not enumerated.

In July, the Bureau will hold an Interagency Conference on Access and Other Barriers to the Statistical Use of Administrative Records. We have invited 13 Federal agencies, staff from the National Academy of Sciences census panels, this Subcommittee's staff, and staff from our Senate oversight subcommittee to participate in this conference. We plan to use the conference to identify barriers to the statistical use of administrative records. This will be an opportunity to discuss our plans for using administrative records in the 1995 test with Federal agencies that have large sets of administrative records.

Work also continues on matching administrative records files, such as voter registration, school district, and tax assessment files, to data collected in the special censuses of Godfrey, Illinois and South Tucson, Arizona. We expect to have the final report for Godfrey very soon and the final report for South Tucson by the end of October 1993.

### *Improving Mail Response Rates*

We are nearing the completion of the data-collection phase of the Mail and Telephone Mode Test. Questionnaires for the test were mailed out in March. This test is a national survey of 20,000 households designed to measure respondent preference for using the mail or the telephone to answer the census form. We want to determine if overall response rates can be improved by offering a choice of response modes.

Preliminary results indicate that offering this choice at the time of the initial mailing of the questionnaire to the respondent does not improve response rates, even though nearly 10 percent of households opted to use the telephone to answer the census.

Offering households the option of responding by telephone during subsequent mail contacts after a respondent has had an opportunity to examine the census form received in the mail and has not yet completed it or returned it by mail does improve response rates slightly. We will combine these findings with other response research results to maximize response to the census.

This July, we will mail out questionnaires for the Appeals and Long Form Experiment. This test will allow us to build on successful elements of our previous questionnaire simplification research in an effort to reduce response rate differences between short forms and long forms. You may recall that, for the 1990 census, that difference was 6 percentage points.

The goals of the Appeals and Long Form Experiment are to determine:

1. The influence of alternative motivational appeals on response rates and questionnaire completeness.
2. The influence of alternative respondent-friendly long form designs on response rates and questionnaire completeness.
3. Whether the degree of emphasis on confidentiality in conjunction with different appeals influences response rates and questionnaire completeness.

Results from this test will be available in time to incorporate them into planning for the 1995 test.

### *Technology*

Census Bureau technology research is looking at ways of providing householders multiple ways of responding to the census. Contractors recently sent us a report assessing various household-based technologies and their viability as data collection modes for 2000. These technologies include telephone interviewing, personal computers, voice recognition, touch-tone data entry, voice recording, and interactive cable television. Based on the report's recommendations, we will continue to conduct research on the use of these technologies and to develop appropriate uses of telephone interviewing, including voice recognition entry.

Contractors will provide an additional assessment of publicly accessible technologies that we might use to count people who may not be able to respond to the census, or who cannot be reached, at home. These technologies include providing electronic kiosks at public places, such as libraries, post offices, and shopping malls; and crafting partnerships with employers to allow employees to answer the census from their place of work using personal computers, telephones, facsimile machines, and so forth.

To make any of these multiple-response modes operationally feasible, we must conduct research on how to quickly and accurately perform computer matches of all census returns, regardless of the response source, to determine if a household or an individual has been counted more than once. This is necessary because the use of multiple-response modes means that many census returns will not have an identi-

cation control number. As you recall, in the 1990 census, we generally restricted census returns to paper returns containing these numbers.

At the March 2 hearing, I noted the review of our system for putting data on the census questionnaires into computer-readable format for speedy processing. We have been looking at promising new commercially available technology that has the ability to scan properly designed questionnaires and electronically interpret most answers. Our focus would be to utilize these off-the-shelf technologies, to the extent possible, and integrate them with a commercially acquired digital imaging system.

### *Sampling Research*

We are conducting extensive research on the use of sampling in the 2000 census. Sampling has tremendous potential to reduce costs and other resource requirements, as well as the potential to improve coverage—both of which remain major goals of our design efforts.

Our sampling research can be characterized as sampling to estimate the size of the population, or sampling for counts, and sampling to estimate content. We are conducting empirical research on sampling to estimate population size in three areas: coverage measurement, a truncated census, and sampling for nonresponse follow-up. In our research on sampling to estimate content, we are looking at matrix sampling, or using multiple “long” forms, to collect sample data. Let me describe what we are doing in each of these four research areas.

*Coverage Measurement.* We are focusing our coverage-measurement research on the concept of a one-number census, which requires integrating the census coverage-measurement program with the census enumeration. “One-number” means there would be only one set of official census numbers, instead of more than one set as there was for the 1990 census. “Integrated” means that we have to integrate the enumeration and the coverage measurement program in time to have state population counts for apportionment purposes by the statutory deadline of December 31, 2000.

The first step in our coverage-measurement research has been to identify eight potential methodologies for integrating coverage measurement into the census, and criteria for evaluating these coverage methodologies.

The potential methodologies were then reviewed by Census Bureau staff, members of the National Academy of Sciences Methods Panel subcommittee on sampling and estimation, and the Technical Committee of the 2000 Task Force. Based on guidance provided by these groups, we are going to explore two new methods that we refer to as “CensusPlus” and “SuperCensus”, in addition to the two methodologies used for the 1990 census—the Post Enumeration Survey (PES) and demographic analysis. We are developing a research plan to define the operational and statistical components of the two new methodologies, and we are developing research plans for the (PES) and demographic analysis.

Both the “CensusPlus” and “SuperCensus” methodologies would involve using special enumeration procedures in a sample of census blocks. We would then develop population estimates for nonsampled areas using modeling and estimation techniques. The major difference between the two methods is that while “CensusPlus” would take place after the completion of a regular enumeration, “SuperCensus” would occur simultaneously with a regular enumeration.

*Truncated Census.* The truncated census feature would involve stopping the enumeration of the population at a specific point in time and then estimating the population size by immediately implementing a coverage-measurement survey. The use of integrated coverage-measurement techniques to produce the final, single set of results is intrinsic to a truncated census design. The point at which the enumeration would be stopped could range from not conducting any nonresponse follow-up to stopping nonresponse follow-up when only the last few difficult cases remain.

We are now conducting an empirical study of truncating the census. This study uses 1990 PES data to simulate truncating the census and is considering three dates for stopping census enumeration. We are preparing estimates of the population at each truncation point and are comparing these estimates to the 1990 PES and the 1990 census count. We will examine estimates and their variances at the national level as well as estimates for demographic groups and subnational areas.

*Sampling for Nonresponse Follow-up.* With sampling for nonresponse follow-up, we would select only a sample of the housing units for which a questionnaire was not returned for subsequent interviewer follow-up.

We are currently reviewing the design for an empirical study to evaluate sampling for nonresponse follow-up in the census. This study will use 1990 census data to simulate sampling for nonresponse follow-up. This research, scheduled to be com-

pleted by September, will offer guidance to us in specifying the sample design and size to use in the 1995 test.

*Matrix Sampling.* We are using 1990 census sample items in our research on sampling for content because actual content needs for 2000 are not yet determined. In particular, we are investigating operational and estimation issues associated with the use of matrix sampling, which I described earlier.

We are in the planning stage of an empirical study to simulate matrix sampling using 1990 census long-form data to estimate population characteristics and cross-tabulations. Census Bureau staff have documented five proposed matrix sampling designs. We will select from these designs for the 1995 test.

### *Coverage Research*

A major goal of the research and development program for the 2000 census is to develop ways to reduce differential undercount and improve coverage overall. Undercoverage of the population occurs when we miss whole households, when we miss persons within households that were counted in the census, and when we miss persons not associated with a household, like those in institutions or the homeless. Our research is addressing the development of improved coverage along all these paths. In addition to looking at special enumeration methods, administrative records, and sampling/and estimation techniques to improve coverage, we are conducting other efforts to come to grips with the problem of undercoverage in the census.

We are conducting research to assess the meaning of certain key concepts that respondents apply in determining household membership. We use terms such as "live," "stay," and "household" in the census in developing household rosters. Where people are highly mobile, connected to multiple households, or tenuously attached to any household, they may have difficulty applying these concepts. In addition, as we were forcefully reminded at the Richmond conference, culturally-specific family and household structures may not mesh well with current census residence concepts.

Another example of research into the coverage issue is the 1993 Coverage Test. The objective of this study is to determine if we can improve overall coverage of mail return households through redesign of the questionnaire and the roster questions. This test, which is looking at different issues than those examined in our Simplified Questionnaire Test, is currently in the design stage. However, at this point, we anticipate using a variation of the 1990 form as a control and developing several alternative forms that incorporate several specific kinds of changes.

First, recall that in 1990, respondents were given residence rules and asked to apply them to their household. We would like to test this approach with several fundamental changes in graphical layout, terminology, and overall format.

Second, in 1990 the respondent was restricted to only identifying whole households with a usual home elsewhere—that is, households where every member had a usual home elsewhere. We will test a form that allows the identification of individuals with usual homes elsewhere.

Finally, another alternative is not to ask the respondent to apply the rules, but rather, to gather sufficient facts from each household so that the Census Bureau can apply the residence rules. These "facts" can be thought of as screener questions to "screen out" nonhousehold members. If we collect a complete listing of persons with attachments to the household, the length of their attachment, and addresses for all other residences, the Census Bureau, not the respondent, can make the determination as to the location, of usual residence.

The basic test methodology for this effort will be to mail out the control and experimental forms and conduct a telephone reinterview with those households that return a questionnaire. We are currently working on the sample design, and currently anticipate using a national probability sample with oversampling of minorities and renters.

The information from this test will help us plan better ways to obtain the census "household roster." This test is targeted for this Fall so that we can determine the form of the roster question that we should use in the 1995 test. A secondary objective is to determine if these design changes have an impact on response rates.

### *Continuous Measurement*

We have begun developing and evaluating a prototype design for a continuous measurement system. By continuous measurement we mean collecting minimal content in the decennial census year and providing for ongoing data collection for additional content throughout the decade. With the prototype we are developing, we will be better able to understand the content, estimation, cost, and policy issues associated with continuous measurement designs. Our working assumption in developing this particular prototype is that it would be able to replace data normally provided

by the conventional long form questionnaire. Beyond that, we would envision a larger system incorporating administrative records, statistical modeling, and data requirements now collected in current demographic surveys.

#### STATUS OF CENSUS BUREAU/U.S. POSTAL SERVICE JOINT COOPERATIVE WORK

Since my testimony before you on March 2, several things have occurred related to our cooperative efforts with the U.S. Postal Service.

First, staff from our two agencies have met several times, most recently on March 26, 1993. We have agreed to conduct a test study in which the U.S. Postal Service will perform an automated match of selected addresses from our 1990 census Address Control File against address information they have compiled. Our agencies have identified the areas for testing, and the Census Bureau has provided technical specifications for the address file that we will send to the U.S. Postal Service. To protect the confidentiality of our address files, we will swear in U.S. Postal Service employees and personnel under contract who will have access to our addresses.

As a product of the automated match, the U.S. Postal Service will provide to the Bureau as much information as they can about deficiencies in our files within the constraints of Title 39. We still must develop a process to receive individual address updates within the constraints of Title 39; legislation would be required to allow the U.S. Postal Service to share information about individual addresses.

Second, the staff of this Subcommittee convened a meeting of U.S. Postal Service and Census Bureau officials on April 19 to discuss progress on the cooperative project. We were encouraged by that meeting and hope that the Subcommittees' staff was reassured that the two agencies should be able to work together to develop a strategy for updating our address list and the TIGER data base. Specifically, the U.S. Postal Service may be able to provide sketch maps showing streets and other features for new developments. We could then use this information to update the TIGER data base. Mr. Chairman, we appreciate this Subcommittee's staff support for this program.

Finally, the interagency Joint Committee for Census Planning established by the Census Bureau and the U.S. Postal Service in 1990 met on March 3 to discuss sharing of address information, identification of vacant units, and rural addressing issues. The Joint Committee has established two subcommittees to explore specific areas of cooperation: One is looking at encouraging local efforts to convert from rural-style to city-style address formats; the second is looking into possible changes in legislation that would allow the U.S. Postal Service and/or the Census Bureau to share address information. Both subcommittees and the full Joint Committee meet on a regular basis and are making progress.

#### CLOSING

Mr. Chairman, as you can see, we are conducting an extensive research and development program for the 2000 census. The President's proposed fiscal year 1994 budget provides adequate funding to carry out this work, and to carry forward decennial operational programs and planning for the 1995 test.

That concludes my testimony. I will be happy to answer any questions.

#### RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY CONGRESSMAN SAWYER TO HARRY A. SCARR

*Question 1a.* Please respond to several concerns GAO raises in its testimony: The Bureau is at the same place it was in 1991 with its list of census design components;

Answer. The GAO (and more recently, the House Appropriations Subcommittee) seems to believe that because we have eliminated all 14 design alternatives, we are "back to square one" in identifying fundamental changes to the 1990 design. We disagree with this conclusion for at least two reasons:

- The designs were never meant to be more than a mechanism for studying different ways to combine new ideas (options) for implementing various components of a census. The options, not the designs, are the key. Now that we have completed sufficient technical research and policy discussions to focus on the most promising options for 2000, we do not need 14 different designs to convey the range of alternatives under consideration. Instead, as the Subcommittee, GAO, and others urged, we can move forward with a smaller, but more auspicious, set of options that can be used as "building blocks" to construct one more design—that for the 2000 census. Over the last 3 years, Census Bureau technical research and policy discussions have identified real improvements for the next census in several areas, including researching alternative mechanisms to collect some or most of the data so that the census can focus on accurate counting; identifying methods to increase the use of



sampling and estimation; identifying key barriers to the accurate enumeration of population groups and areas that often are undercounted in the census, and researching methods to overcome those barriers; making a commitment to ensuring corrections for the undercount are folded into the official census numbers; and finding ways to simplify the questionnaires and otherwise improve mail response rates.

• We have spent this time very productively with a wide range of stakeholders on the review and discussion of new ideas, the design alternatives, and criteria for identifying the most promising ideas—what we call our “open process”—and that effort has been invaluable. They, not the Census Bureau, identified the new ideas, the 14 design alternatives, and decision criteria. This open process has allowed everyone involved in this effort, both us and our customers, to better understand both common and divergent goals, and to surface those policy, legal, and technical issues that will be key to the final choice of a design.

*Question b.* The Administration did not request sufficient funds in fiscal year 1994 for several critical activities, including data processing technology and the 1995 test census.

Answer. We believe we had sufficient funding in the Administration's 1994 request to carry out current plans for technology and the test census. Current plans include the testing of one set of the most promising options for the next census in four sites; earlier plans included the testing of as many as four separate designs in eight sites.

In the area of technology, we plan to complete the Data Capture System 2000 in 1996, because its completion in 1995 would have required greatly accelerated resources and funding.

We will need to reexamine all these plans, or set new priorities, to the extent the final FY 1994 appropriation differs from our request.

*Question c.* The Bureau still does not know how it will prepare the geographic components of the TIGER system in time for the next census.

Answer. The Census Bureau has included in the President's FY 1994 budget request its proposal for a cooperative project with the United States Postal Service (USPS) to create and continuously update a Master Address File (MAF) linked to the TIGER data base. This proposal outlines a comprehensive plan and methodology for updating the road and address information in the TIGER data base. Assuming that the Congress approves the requested funding, we will begin full-scale operations to implement this continuous process in FY 1995. These activities will provide an up-to-date geographic base for the 2000 decennial census as well as for the Census Bureau's other statistical programs.

The methodology calls for the Census Bureau and the USPS to focus their initial activities on the task of supporting the most critical, basic requirements for the 2000 decennial census: creating a complete and accurate list of housing unit addresses. The two agencies will do this by periodically matching the MAF addresses (which are linked to the TIGER data base) to the USPS delivery sequence file of addresses. Based on this matching process, the two agencies can identify addresses missing from the MAF and streets that are missing from the TIGER data base. As part of the address list sharing test that the two agencies are conducting to determine the procedures that will be used to conduct the full-scale cooperative project, USPS staff will sketch the locations of streets determined to be missing from the TIGER data base. The USPS staff will do this on maps provided by the Census Bureau to facilitate their addition to the TIGER data base.

In addition to its proposed partnership with the USPS, the Census Bureau is working actively with its longstanding partner, the U.S. Geological Survey, as well as state and local governments across the United States, to develop methodologies for the interchange of updated geographic information. This work is taking place under the auspices of the Federal Geographic Data Committee (FGDC) as part of the Census Bureau's participation in the broader effort to coordinate mapping activities across all Federal agencies and among the various levels of government and private sector organization in the United States. We believe that this combination of USPS and FGDC activities will provide not only an appropriately updated TIGER data base, but also a current, linked address list to support the 2000 decennial census and the needs of the Census Bureau's other major statistical programs.

*Question 2.* The Academy has identified a major policy choice for future censuses: a census that tries to count every person through direct or indirect contact, and a census that emphasizes the mail phase, but then captures nonrespondents through sampling or other statistical techniques.

Do you envision comparing these 2 basic designs during the 1995 test census, so that Congress can make an informed choice about which type of census it wants to fund?

Answer. Much of this comparative information can be simulated using 1990 data files and cost modeling. This work is underway as part of the research on sampling for nonresponse follow-up that we will complete by this September. We also plan to use sampling for nonresponse in the 1995 Census Test so that we can gain operational experience and additional data for cost, coverage, and data quality analyses.

*Question 3.* The Bureau has just completed research on nontraditional ways of responding to the census, such as personal computers and facsimile machines. If those are viable options, why was the Multi-Mode Test limited to the mail and telephone?

Answer. The Census Bureau conducted research on the use of the telephone as an alternative mode to mailing back a completed questionnaire based on the assumption that offering a variety of modes to the public may increase their participation in the census (that is to say, it may have some positive impact on response rates). We chose the telephone for the first test of alternatives because it is the most universally available mode, next to the mail. We believe that any additional gains in census participation attributed to other response modes such as personal computers or FAX machines would be less than the effects from offering the telephone as an option.

Preliminary evidence from the mail and telephone mode test indicates that there is no improvement in response rates as a result of offering a choice of modes (mail or telephone) to respondents. Evaluation of data quality differences between telephone and mail response to the census is not yet complete.

The Mail and Telephone Mode Test tells us that at a national level, about 5.5 percent of households prefer to answer the census by telephone. However, these same households would have completed the census return by mail if the telephone option was not made available to them.

Even if offering additional response modes does not increase response rates, we will continue to consider the development of additional response modes such as the personal computer and FAX machine as it is clear that some households prefer response modes other than the mail. A secondary benefit of these alternative electronic response modes is that the data are automatically captured and edited at the time the respondent provides the data, eliminating a costly follow-up contact if mail-return data are found to be incomplete or in error after they are captured and edited.

*Question 4.* Your testimony mentioned the testing of "matrix sampling" in your 2000 research program. How will the Bureau determine the content of each sample form for the 1995 test?

Answer. In September, based on research now underway, we will determine a specific matrix design to be used for the 1995 Census Test. The test will provide operational experience and additional data for cost, coverage, and data quality analyses of matrix sampling.

The key research question now is to identify the best overall approach for dividing topics over multiple sample forms. Besides determining the "right" number of forms, this also includes determining the extent to which there should be a common "core" of questions on all forms and determining the best way to estimate (or model) cross-tabulations of items not collected on the same form.

*Question 5.* At a March 5, 1993 meeting of the Commerce Department Advisory Committee on the 2000 Census, Dr. Robert Tortora said that the Census Bureau was developing a policy paper on local government involvement in the 2000 census. What is the status of that effort?

The Census Bureau is actively researching possible interactions with local governments for the 2000 census. Among the types of activities being considered, where local governments might be involved, are:

- Providing governmental unit address lists (such as assessment records and the like) for creating and updating the MAF;
- Providing governmental administrative records (such as school records, voter registration records and the like) to augment enumeration data; and
- Providing staff and space for assistance centers, outreach and promotional support, and recruitment.

The Census Bureau is researching the legal and policy implications of sharing its address list with state and local governments. We are testing in special censuses the accessibility, coverage, and content of governmental administrative records. We have had experience in working with local governments in support of outreach and promotion through Complete Count Committees and the like in 1990, but would like to explore the possibility of contractual agreements in this area.

At this time, Field Division is conducting a series of meetings with a number of local governments in its twelve regions to document issues of concern of the local governments and how we might better keep them informed of census developments.

This information will be used to shape an informational program for local governments.

At the conclusion of this research agenda, the findings will be documented and circulated with recommendations for local government interactions.

*Question 6.* What efforts are currently underway to develop improved methods of counting the homeless?

*Answer.* We are examining three specific approaches at this time in regard to the homeless enumeration. First, we will look at a possible alternative method of enumeration using sampling and/or estimation techniques as part of the method. There are legal implications, however, because of the use of sampling.

Second, we are looking at what local areas have done to count their homeless populations since the 1990 census. By examining these local approaches, we might find a more operationally efficient and complete enumeration methodology.

Third, we will examine a daytime enumeration method that relies heavily on service providers for enumerating the homeless. This involves enumerating homeless persons where they receive services.

We are just beginning to investigate and test these approaches.

Mr. SAWYER. Thank you. Thank you very much.

The Academy made their three specific recommendations: the importance of doing a lot of research on sampling techniques for nonresponse follow-up, the use of the post-enumeration survey to determine accuracy of data at small geographic levels, and the question of whether the Bureau should initiate work to develop a separate, long-term research program leading toward an independent administrative records census.

Could you respond to those three for me?

Mr. SCARR. Certainly, Mr. Chairman. We agree completely with the panel's first recommendation that serious consideration be given to investigating sampling for nonresponse follow-up. Indeed, we will be doing that kind of research in 1995, so we are in complete agreement with the Academy on that.

With respect to their second recommendation, quite candidly, we are working with them now to determine exactly what sort of analysis they want of the 1990 data, but we have no problem with that. We are considering that.

Mr. SAWYER. You can play with it endlessly.

Mr. SCARR. That is about the only problem we have. Aside from that, as I say, we are working with them to find out exactly what they want to do.

Their third recommendation is more complex and our response is a little more complex: The recommendation is to initiate a separate program of research on uses of administrative records not directly related to the 2000 census focusing primarily on the 2010 census and on current estimates programs.

We are considering that at the moment and we will work with them to understand exactly what they mean and we will move forward with something like that.

Mr. SAWYER. Let me suggest that, from a personal point of view, it would seem to me that that might not require only a substantial change in what we mean by administrative records, but possibly a constitutional change since it really implies a very different kind of instrument.

Mr. SCARR. Yes, that is why we have some problems with it.

Mr. SAWYER. From my point of view, while I tend to believe that administrative records hold enormous promise, I just really had not ever envisioned a wholly administrative records census. The Academy concluded similarly, though, that sampling for nonresponse

follow-up and for coverage improvement, at least by their constitutional scientists, was constitutionally and statutorily feasible.

Is that your view, or do you think that we still need specific legislation for the use of sampling for those limited purposes?

Mr. SCARR. Mr. Chairman, we have from the Congressional Research Service the opinion that that would be permissible, and we now have from the Academy that that would be permissible. I continue to believe that any guidance that the Congress could offer would be welcome because, quite candidly, I don't see it as clearly as they do.

Mr. SAWYER. I understand. It is probably true, as you suggest, that 1995 would be a very difficult arena in which to test continuous measurement techniques. Have you given thought to how you might go about doing that, other than 1995?

Mr. SCARR. There are several efforts underway. First, the zero year part of a continuous measurement census would very likely be what we have described as a minimal Voting Rights Act Census, so everything in connection with that will be done in the 1995 test. As the research proceeds to the—

Mr. SAWYER. I assume that you are viewing that as an extreme definition, that there are other definitions.

Mr. SCARR. There are other definitions, yes, that is an extreme definition. But whatever the year zero census is likely to be and would be likely to be in a continuous measurement census, I am confident that we will have tested that in the 1995 test.

I think that the issues that Mr. Schultze raised are the critical issues, an issue of cooperation with other Federal agencies, being sure that they are comfortable with what is being done, the issue of cost.

From the standpoint of empirical testing, we would hope if a design like that were to be the design that were chosen, we would expect to begin doing something with a larger frame sample in 1997 or 1998 at the latest, probably during those 2 years, in order to see if it was feasible. Because if it wasn't feasible, if it didn't work and if it fell apart, you would want to be in a position to be able to conduct a successful year 2000 census using some other design.

But that is at the moment, Mr. Chairman, that is really a guess, but that is sort of the time frame and the framework and the way we are looking at it.

Mr. SAWYER. And clearly we are probably looking beyond 2000 for application in any case.

Mr. SCARR. Yes. What I was suggesting is, if there were to be a continuous measurement census and the 2000—and sort of starting with the 2000 census, in quotation marks, you really have to start beforehand. And that is what—

Mr. SAWYER. It is what we mean when we talk about getting beyond the 10-year horizon.

Mr. SCARR. Yes, we have to look at that very carefully. Because as I indicated, there are real trade-offs.

Mr. SAWYER. In 1995, have you decided which questionnaires you are going to test? Are you still planning to use the 1990 forms, the short and the long form?



Mr. SCARR. We plan to use 1990 content questions, Mr. Chairman, but we will use them in a user-friendly form in most of the test instances.

Mr. SAWYER. So you are not going to use those as they were even as a control.

Mr. SCARR. It is my understanding, no, no. We have committed—at least in our judgment, there has been that much fundamental change already.

Mr. SAWYER. The Bureau altered its design selection criteria, saying that the Census, I quote, "should provide small area data that the Census is uniquely capable of supplying."

Could you expand for us the notion of small area, what your working definition is and how you will determine the level of geography needed for specific data?

Mr. SCARR. If I could treat that—let me treat that in two parts. Let me answer one part, and then let me ask Susan if she would deal with the second part. The first part is the change in that criterion was basically to try and clarify what we meant by it. There is nothing more to it. It was to make it clear that what we meant was the Census should provide data that the Census can uniquely provide and nothing more.

With respect to determining the level of disaggregation and how they will determine that, there are a variety of factors that enter into it, and Susan is better prepared than I am to answer that.

Ms. MISKURA. For our purposes, we would consider a small area, particularly with regard to the collection of sample type data, to be the kinds of small areas we would traditionally use. For the Census, the smallest area for which we would do sample data is the tract number or block group. And there are also lots of different kinds of geographic areas like school districts that could be small in some sense.

Mr. SAWYER. Something of some considerable ongoing concern for me and a number of folks I serve with on another committee.

Ms. MISKURA. Yes. And accumulating blocks or accumulating tract data to school districts is an extremely important applications and there are others. So I think the definition of small area would be kind of what we have done or thought of in the past.

And our target for being able to produce good characteristics for small areas would be sort of the benchmark, would be how good are they compared to what they have been in the past. So that even for our work on a continuous measurement system, we really want to see if we can produce, although it may be accumulated data over time, data that is accurate down to the tract level.

Mr. SAWYER. Can you comment on Dr. Schultze's remark with regard to the expectations for accuracy in those small areas as opposed to counterbalancing inaccuracies as they may be aggregated?

Ms. MISKURA. I think Dr. Schultze and the panel very perceptively diagnosed the situation, that in the past, very small block level data have been available. Users think about these bigger small areas as being accumulations of blocks. And probably a more productive mind-set is to think about the areas themselves and think about tracts, think about school districts and traffic zones those are the things that are actually used. Those are the areas that problematic or funding decisions are made on.

So the real key may not be the error at the block level or the total error at the block level, but really at these accumulated areas. And I think that is what their recommendation gets to.

Mr. SAWYER. And that aggregate error may be a more meaningful number than the cumulative error, which may have no meaning at all?

Ms. MISKURA. Right.

Mr. SAWYER. OK. I just wanted to make sure I understood the concept. I am getting very worried about time and the afternoon that we are in. I have several other questions, but if you would be comfortable answering those in writing as opposed to spoken testimony today, just let me say thank you for your testimony today.

Mr. SCARR. Thank you, Mr. Chairman.

Mr. SAWYER. It's a pleasure to work with you.

Our final witness this afternoon is William M. Hunt, who is the Director, Federal Management Issues, General Government Division, U.S. General Accounting Office. Let me say welcome.

Mr. HUNT. Good to see you again, Mr. Chairman.

Mr. SAWYER. Good to see you. It probably won't be the last time. As always, your full testimony will be entered as part of the record and if you could identify your colleagues for that record, I am sure that the stenographer would appreciate it.

**STATEMENT OF WILLIAM M. HUNT, DIRECTOR, FEDERAL MANAGEMENT ISSUES, GENERAL GOVERNMENT DIVISION, U.S. GENERAL ACCOUNTING OFFICE; ACCOMPANIED BY BRUCE JOHNSON, ASSISTANT DIRECTOR, AND JACK KAUFMAN, SENIOR EVALUATOR**

Mr. HUNT. Thank you, Mr. Chairman. I am pleased to be here today. To my right is Bruce Johnson who oversees our census and other statistical policy work at GAO, and to my left is Jack Kaufman who has been GAO's focal point and in-house expert on census issues for many years.

As you recall, we discussed in June 1992 that a key measure of continued census redesign progress would be the rate at which the Census Bureau rejected design alternatives to enable it to concentrate on the most promising ones for the 2000 census. At that time, we said that if the full spectrum of 14 alternatives was still being considered 1 year later, the possibility for meaningful change for the 2000 census would be severely diminished.

I testified in March of this year that a lack of Bureau progress in redesigning the 2000 census jeopardized the prospects of fundamental reform. Since then, the Bureau has altered its decision-making approach and taken steps to refocus its census research and development efforts. It has rejected all 14 design alternatives that were under study for over a year. Instead, it is now concentrating on integrating promising features into a new design for possible application in the 2000 census.

With the elimination of all 14 design alternatives, the Bureau now faces the formidable task of deciding which features will be combined into an integrated design to be tested in 1995. This is essentially the same position the Bureau was in when it began the 2000 census redesign effort over 2 years ago. Given the added time

constraints now facing the Bureau, it must quickly identify the most promising features and prepare operating plans for the 1995 test based on a new basic design that is yet to be determined.

In the 4 months remaining until the end of September, the Bureau must assess whether the features under consideration comply with legal provisions, evaluate them against its decision-making criteria, and determine how these features can be integrated operationally. If features are not tested in 1995, it is unlikely that they will be included in the 2000 census.

While the Bureau's recent efforts cannot make up for valuable lost time and resources, fundamental breakthroughs in census taking are still possible for the 2000 census. Such breakthroughs, however, will require prompt consensus, policy guidance, and leadership from the Office of Management and Budget, and senior Department of Commerce and Bureau officials. These officials need to build consensus with the Congress and key stakeholders as well as gain the public support necessary for major census innovations.

Overall, we believe that the 2000 census must strike a reasonable balance among coverage accuracy, cost considerations and Federal content needs. But for over 200 years, a primary criterion by which to judge the success of the census has been and continues to be coverage accuracy. That is enumerating all persons and enumerating them at their proper location.

Acknowledging the impossibility of achieving a perfect count, the Bureau in recent censuses has modified the coverage accuracy goal to include reducing the differential undercount, whereby the population of racial or ethnic groups are not disproportionately missed in the census.

In the austere budget environment facing the Federal Government in this decade, containing the cost of the next census must unquestionably be a primary criterion. If cost efficiencies are not vigorously pursued as part of the 2000 census design, there is a risk that the 2000 census will not be completed as planned.

In the past two censuses, costs have been miscalculated and the Bureau has needed last-minute emergency supplemental appropriations to complete them. If this need arises again in the 2000 census, given the bleak fiscal condition facing the Federal Government, there is no assurance that the Congress will provide the necessary financial bailout.

Under such a scenario, the 2000 census may be truncated by financial necessity rather than through prudent planning, which will pose unknown risks for data quality.

Mr. SAWYER. Kind of unplanned reform.

Mr. HUNT. Yes, sir.

The third primary criterion is collecting sufficient data to satisfy Federal program needs. The census currently is the only source that provides needed data for small areas to redraw legislative and other political boundaries, to enforce the Voting Rights Act, and to allocate Federal funds such as education funds targeted to school districts, housing funds targeted to census tracts, and/or transportation funds targeted to local governments.

In meeting these Federal program needs for small area data, the census must strike a balance and avoid placing such an undue burden on respondents that meeting data needs would unaccept-

ably interfere with improving coverage accuracy or controlling costs. A close scrutiny of Federal data needs is crucial to census planning efforts.

In December 1992, OMB asked Federal agencies to identify their data needs from the 2000 census. In its request, OMB also specifically asked the agencies to identify possible alternative data sources and to comment on two design alternatives under consideration by the Bureau that would affect data content, that is expanded content and continuous measurements.

The Bureau is currently reviewing the Federal agency's responses. Preliminary information shows that the agencies want the same type and amount of data as collected in the 1990 census. In general, Federal agencies did not identify alternative sources to satisfy their data needs. Also, many had reservations about other design alternatives that featured changes in data availability or reliability.

The Bureau and OMB officials need to scrutinize those responses to determine what the Bureau must do to meet Federal data needs. The challenge is to weigh the benefits of these data against the cost of collecting them. The fundamental issue underlying the census redesign effort is whether the Census can and should continue to be the only source for much of the Nation's population and housing information.

The National Academy of Sciences panel, whose chairman has testified here today, will assist in the systematic review of population and housing data needs. The Academy's panel is examining the need for data at different frequencies and geographic levels of detail. It also is examining the availability and cost of alternative methods to obtain these data. We urge officials in the executive branch to examine carefully how to meet Federal data needs for some national data, by a once-in-a-decade census, by more frequent sample surveys that produce less geographically detailed data, or by administrative records.

There is a desire to obtain more socio-demographic data in-between censuses for some national areas. For example, legislation was passed in 1976 mandating a mid-decade census. The Congress hoped that the cost of conducting such a census would generally be offset by reductions in other statistical series. Because the Bureau could not identify substantial cost offsets, OMB did not request and Congress did not provide, funds to take a mid-decade census.

We believe, given the likely cost of expanding the collection of data between censuses, the only feasible offsets of sufficient magnitude are reductions in the cost of the decennial census. Cost efficiencies appear most probable for the 2000 census by using procedures to obtain greater public cooperation and to use sampling during follow-up efforts.

Moreover, the availability of continually updated address files and geographic data developed primarily for the decennial census, would enhance the accuracy and reduce the cost of intercensal data.

One of the best ways to contain the costs of the census is through prudent investments today that can lead to savings tomorrow. We are concerned that the fiscal year 1994 budget request has not al-



lowed sufficient funds for some important research and testing activities.

The adequacy of census planning in the next 2 years will have significant implications for the 2000 census and beyond. And the availability of sufficient funding is important for that planning. Specifically, we question whether the Bureau's 1994 budget request contains sufficient funding for data capture research, improvements in the Bureau's address list and automated geographic system, and 1995 test preparations.

My written testimony provides more details on these concerns.

In closing, let me summarize the situation facing us and what must be done. The Bureau's efforts to assess design alternatives have consumed valuable time and scarce resources. In the time remaining before the September 1993 deadline, the Bureau, the Department of Commerce, and OMB must quickly focus on those features viable for the 2000 census that offer genuine opportunities to improve the coverage accuracy of the census at a reasonable cost, while meeting Federal data needs.

In my written testimony, we have identified several opportunities that should be considered for testing in 1995. These opportunities include: One, improving public cooperation through the use of more user-friendly, streamlined census questionnaires and more frequent mail contacts; two, reducing costly and error-prone field follow-up efforts by using sampling techniques to follow up with households who do not mail back their questionnaires; three, obtaining an improved address list linked to the Bureau's geographic system, primarily by greater reliance on the postal service; and four, developing an approach for a one-number census.

While the Bureau also recognizes these as valuable opportunities, much more needs to be done and time is running out. We urge the Bureau to take decisive action to ensure that these opportunities are thoroughly explored. As our society has changed in fundamental ways, so must fundamental changes occur in the census. Such census breakthroughs can only occur if they are properly planned, tested and implemented.

Bureau, Commerce, and OMB officials collectively and cooperatively must assume the responsibility for leading this effort. They must clarify the necessity and purpose of fundamental change, establish the proper sense of urgency, develop plans for its implementation, and devote appropriate resources to its accomplishment.

Mr. Chairman, this concludes my formal statement. My colleagues and I would be pleased to answer any questions that you might have.

[The prepared statement of Mr. Hunt follows:]

PREPARED STATEMENT OF WILLIAM M. HUNT, DIRECTOR, FEDERAL MANAGEMENT  
ISSUES, GENERAL ACCOUNTING OFFICE

SUMMARY STATEMENT

Since GAO's testimony last March on 2000 census planning, the Census Bureau has altered its decisionmaking approach and taken steps to refocus its research and development efforts. Driven by the Bureau's impending September 1993 deadline for deciding which designs to test in 1995, the Bureau has recommended eliminating all 14 design alternatives that have formed the framework of its research and consultation program for the last year.

Little time is left to achieve fundamental breakthroughs. In the next four months, the Bureau must develop a new basic design to test in 1995. While the Bureau's recent efforts cannot make up for valuable lost time and resources, fundamental breakthroughs in census taking are still possible for the 2000 census.

Such breakthroughs can occur only if they are properly planned, tested, and implemented. Bureau, Commerce, and Office of Management and Budget (OMB) officials must clarify the necessity and purpose of fundamental change, establish the proper sense of urgency, develop plans for its implementation, and devote appropriate resources to its accomplishment. These officials also need to allow sufficient time to build consensus with Congress and key stakeholders, as well as gain the public support necessary for major census innovations.

GAO believes the Bureau's set of decisionmaking criteria is too long and not sufficiently focused on what is most important—improving coverage accuracy, containing costs, and meeting federal data needs. Opportunities still exist to meet this more focused set of criteria. GAO continues to urge the Bureau to focus on simplifying and streamlining the census questionnaire, sampling during certain costly field operations, increasing cooperation with the U.S. Postal Service in building an accurate address list, and developing an approach for taking a "one-number" census, as opposed to the 1990 census that produced two sets of numbers.

Bureau and OMB officials need to scrutinize data requested by federal agencies to determine what the Bureau must do to meet their needs. The content of the census questionnaire and two possible census breakthroughs—matrix sampling and continuous measurement—depend upon this close examination.

The 1994 budget requested by the administration for census planning may underfund some important activities. Specifically, GAO is concerned that data processing research and development, geographic support activities, and test census preparations may be underfunded.

---

Mr. Chairman, Mr. Petri, and Members of the Subcommittee:

I am pleased to be here today to discuss the Census Bureau's progress planning the 2000 Decennial Census. You requested our assessment of the Census Bureau's progress toward its key September 1993 deadline for establishing 1995 test census objectives. Specifically, you asked us to discuss the development and use of criteria for determining what to test in 1995; Bureau progress in analyzing data needs and possible content of the 2000 census; and the adequacy of the recently submitted fiscal year 1994 budget request as it relates to 2000 census preparation. My testimony is based on our continuing audit work, at the Subcommittee's request, to monitor and evaluate 2000 census planning activities and operations.

We testified in March, 1993<sup>1</sup> that a lack of Bureau progress in redesigning the 2000 census jeopardized the prospects of fundamental reform. Since then, the Bureau has altered its decisionmaking approach and taken steps to refocus its census research and development efforts. It has rejected all 14 design alternatives (each composed of different operational features) that were under study for over a year. Instead, it is concentrating on integrating promising features into a new design for possible application in the 2000 census. This action essentially reverts to the Bureau's earlier approach, which was previously published in March 1991.

Now in the 4 months remaining until the end of September, the Bureau must intensively research and refine the design features that appear most promising. It must develop a new basic design to test in 1995 by assessing whether the features under consideration comply with legal provisions, evaluating them against its decisionmaking criteria, and determining how these features can be integrated operationally.

While the Bureau's recent efforts cannot make up for valuable lost time and resources, fundamental breakthroughs in census taking are still possible for the 2000 census. Such breakthroughs, however, will require prompt consensus, policy guidance, and leadership from the Office of Management and Budget (OMB) and senior Department of Commerce and Bureau officials. These officials need to build consensus with the Congress and key stakeholders, as well as gain the public support necessary for major census innovations. They also need to test and integrate the many interwoven procedures and technologies that constitute an undertaking as big and as complex as the decennial census.

---

<sup>1</sup> See *Decennial Census: Fundamental Reform Jeopardized by Lack of Progress* (GAO/T-GGD-93-6, Mar. 2, 1993)

## THE BUREAU HAS ADOPTED A NEW DECISIONMAKING APPROACH

Driven by its impending September 1993 deadline for deciding which designs to test in 1995, earlier this month the Bureau adopted a new decisionmaking approach. Based on the Bureau's analysis of the 14 census design alternatives, it has recommended eliminating all of the alternatives because they did not adequately meet the criteria the Bureau established. These design alternatives had been the framework of its research and consultation program for over 1 year. Instead, it is now redirecting its research efforts toward selecting the most promising features of those design alternatives for development and testing in 1995.

*Dates for Design Alternative Recommendations Moved Up*

The Bureau recently revised its schedule for completing its analysis of each design alternative by issuing design alternative recommendations (DAR). Formerly not scheduled for completion until August 30, 1993, all of the DARs were completed in May 1993. In part, this expedited schedule responds to concerns we expressed almost 1 year ago about the Bureau's lack of progress in deciding on alternative census designs. We testified in June 1992 that a key measure of continued census redesign progress would be the rate at which the Bureau rejected design alternatives to enable it to concentrate on the most promising ones for the 2000 census.<sup>2</sup> At that time, we said that if the full spectrum of 14 alternatives was still being considered 1 year later, the possibility for meaningful change for the 2000 census would be severely diminished.

*All Design Alternatives Recommended for Elimination*

We have not had the opportunity to study in detail the final DARs and all of the supporting documentation which led the Bureau to recommend elimination of all 14 design alternatives. As we have said in the past, several of the design alternatives appeared from the outset not to comply with the Constitution or existing legislation and thus were incompatible with what became the Bureau's mandatory criteria. For example, one design alternative—sampling for the count—appears to violate the Constitutional requirement that an "enumeration" of the entire population be made for the apportionment of representatives. Valuable time and resources were invested in setting up and justifying the elimination of such improbable and less likely alternatives.

Despite the Bureau's recommendation to eliminate all 14 design alternatives, the Bureau still faces the formidable task of deciding which features will be combined into an integrated design to be tested in 1995. This is essentially the same position the Bureau was in when it began the 2000 census redesign effort over 2 years ago. Given the time constraints now facing the Bureau, it quickly must identify the most promising features and prepare operating plans for the 1995 test based on a new basic design that is yet to be determined. If features are not tested in 1995, it is unlikely that they will be included in the 2000 census.

*Request for Public Comment Was Reduced in Scope*

The Bureau also changed its approach for soliciting public comments on its decisionmaking process. The Bureau notice in the Federal Register requesting public comment on its decisionmaking criteria was modified significantly from an earlier draft. The final Federal Register notice only requested comment on the criteria used to evaluate the design alternatives. It did not request comments on the DAR process or promise to publish completed DARs in the Federal Register for notice and comment.

We always have encouraged and supported the Bureau's desire to open the census planning process to others. In our March 1993 testimony we said, however, that the Bureau's process for identifying promising census designs was time-consuming and cumbersome and that it might impair the Bureau's ability to institute major innovations for the 2000 census. In part, the changes in the Federal Register notice responded to our concern that the Bureau needed to simplify and streamline this process.

## BUREAU SLIGHTLY MODIFIED CRITERIA IN RESPONSE TO PUBLIC COMMENTS

The Bureau's final evaluations of the 14 design alternatives were based on 16 criteria. Originally, the Bureau proposed 18 criteria; 5 were considered mandatory because they represent legal requirements, and 13 were considered desirable. After re-

<sup>2</sup> See *Census Reform: Major Expansion in Use of Administrative Records for 2000 is Doubtful* (GAO/T-GGD-92-54, June 26, 1992).

ceiving public comments through the Federal Register notice, the Bureau retained all of the mandatory criteria, switched one criterion—the ability to reduce the differential undercount—from the desirable to the mandatory category, eliminated two of the desirable criteria,<sup>3</sup> and made several minor language changes to other criteria.

#### BUREAU NEEDS NEW, MORE FOCUSED SET OF CRITERIA TO GUIDE FUTURE DECISIONMAKING

Now that the Bureau has recommended eliminating all 14 design alternatives, it is not clear whether and how the Bureau intends to continue using the 16 criteria for future census planning. For guiding future decisionmaking, particularly its choice of features to test in 1995, we believe the Bureau should use a new, more focused set of criteria. The set of 16 criteria used for evaluating design alternatives is too long and not sufficiently focused on what is most important. The division of the criteria into mandatory and desirable categories does not succinctly capture the essence of what a redesigned census needs to accomplish. Five of the six mandatory criteria essentially do no more than recognize that the Bureau must meet the requirements of the Constitution and satisfy other statutory obligations or requirements for data. At some point soon, the Bureau and others in the executive branch must agree on a more manageable and meaningful list of the most important desired outcome-oriented criteria to ensure that the Bureau's research, development, and testing activities focus on what truly needs to be accomplished in the 2000 census.

In his testimony on March 2, 1993, the Acting Director of the Bureau acknowledged that legal requirements for apportionment counts, state redistricting data, and age and race/ethnicity data that are required to enforce the Voting Rights Act must be satisfied. In addition, he said that the following "three overarching concerns" would guide the Bureau: (1) reducing the differential undercount and making the census more accurate overall, (2) containing costs, and (3) keeping the process open.

Mr. Chairman, your April 9, 1993, letter to the Bureau responding to the Federal Register notice also expressed concern that the large number of criteria might overshadow the chief goals of the census. You suggested a clearer approach might be to identify the primary aims of the next census: You also stated your belief that reducing the differential undercount and containing cost should be the two most important goals. You added that such clarity of purpose would help the Bureau communicate its mission to others, as well as provide a focus for its own work. We strongly agree. Overall, we believe that the 2000 census must strike a reasonable balance among coverage accuracy, cost considerations, and federal content needs.

For over 200 years, a primary criterion by which to judge the success of the census has been and continues to be coverage accuracy—enumerating all persons and enumerating them at their proper location. Acknowledging the impossibility of achieving a perfect count, the Bureau, in recent censuses, has modified the coverage accuracy goal to include reducing the differential undercount, whereby the population of racial or ethnic groups are not disproportionately missed in the census.

In the austere budget environment facing the federal government in this decade, containing the cost of the next census must unquestionably be a primary criterion. If cost efficiencies are not vigorously pursued as part of the 2000 census design, there is a risk that the 2000 census will not be completed as planned. In the past two censuses, costs have escalated, and the Bureau has needed last-minute emergency supplemental appropriation to complete them. If this need arises again in the 2000 census, given the bleak fiscal condition facing the federal government, there is no assurance that the Congress will provide the necessary financial bailout. Under such a scenario, the 2000 census may be truncated by financial necessity rather than through prudent planning—posing unknown risks for data quality.

The third primary criterion is collecting sufficient data to satisfy federal program needs. The census currently is the only source that provides needed data for small areas to redraw legislative and other political boundaries, to enforce the Voting Rights Act, and to allocate federal funds such as education funds targeted to school districts, housing funds targeted to census tracts, or transportation funds targeted to local governments. In meeting these federal program needs for small area data, the census must strike a balance and avoid placing such an undue burden on the re-

<sup>3</sup> Eliminated were the following two desirable criteria: (1) methods and procedures are understandable and credible to the public and (2) confidence that related aspects of the federal statistical structure will be in place to support the census.



spondents that meeting data needs would unacceptably interfere with improving coverage accuracy or controlling costs.

#### OPPORTUNITIES EXIST FOR IMPROVING ACCURACY, CONTAINING COSTS, AND MEETING FEDERAL DATA NEEDS

We have advocated for some time that opportunities exist for achieving coverage accuracy and cost efficiencies while still meeting federal data needs and they should be considered for testing in 1995. These opportunities include (1) improving public cooperation through the use of more user-friendly, streamlined census questionnaires and more frequent mail contacts; (2) reducing costly and error-prone field follow-up efforts by using sampling techniques to follow up with households who do not mail back their questionnaires; (3) obtaining an improved address list linked to the Bureau's geographic system primarily by greater reliance on the Postal Service; and (4) developing an approach for a one-number census. While the Bureau also recognizes these as valuable opportunities, much more remains to be done, and time is running out. We urge the Bureau to take decisive action to ensure that these opportunities are thoroughly explored.

#### *Opportunities for Increasing Public Cooperation*

Public cooperation is essential for obtaining accurate census data at a reasonable cost. The Bureau's 1992 research demonstrated the usefulness of simplifying and streamlining the census questionnaire and the effectiveness of multiple mail contacts. In the Bureau's tests, these changes resulted in insignificant improvements in mail response. The Bureau projected that these changes could save about \$0.5 billion (in 1992 dollars). At the April 27, 1993, hearing held by the House Subcommittee on Commerce, Justice, State, the Judiciary, and Related Agencies, Committee on Appropriations on the Bureau's appropriation request for fiscal year 1994, Members, commenting on 2000 census planning efforts, emphasized the importance of counting the population and simplifying the form.

#### *Opportunities for Reducing Costly Field Work*

Even with a simplified questionnaire, the Bureau's 1992 test showed that the Bureau can still expect that a large number of households will not return their forms by mail. Statistical sampling of these nonrespondents would reduce the time and costly labor-intensive fieldwork now required to follow up on missing questionnaires. The Bureau only recently initiated research that could provide a basis for using sampling to follow up nonresponding households. A key objective of the Bureau's research on sampling is to identify the point at which sampling can provide data of comparable quality to the existing follow-up procedures in a more cost-effective manner. The Bureau estimated that it could have saved \$460 million if it had sampled 10 percent of nonresponding households in the 1990 census rather than attempting to collect data on all of them. The Bureau also will be proposing legislation to make it clear that sampling may be used to estimate the number and characteristics of people in nonresponding households.

#### *Opportunities for Improving the Address List*

An accurate and complete address list and associated geographic information is a cornerstone of a successful mail census. For several decades, the Bureau has recognized the valuable contribution that the Postal Service could make to improve the quality of the address list. In the past few years, the Bureau also has recognized the value of maintaining an accurate and complete address list throughout the decade. However, only in the past few months, through the encouragement of this Subcommittee, have the Bureau and the Postal Service initiated a feasibility test of sharing address information on a long-term basis.

Concerns about the legal authority of the Postal Service to share its information with the Bureau still need to be resolved so that the Bureau can take full advantage of the address and occupancy status data now maintained on computerized files by the Postal Service. The importance to a mail census of achieving optimal cooperation and communication between the Bureau and the Postal Service cannot be overstated. Both organizations must give this cooperative effort the attention it deserves.

Despite progress on exchanging address list information, the Bureau has been unable to obtain the hoped-for level of cooperation from the Postal Service for updating the Bureau's automated geographic system. The Bureau proposed sharing its geographic information system data, enhanced to meet special Postal Service needs, with the Postal Service. In return, the Postal Service would assist in updating the Bureau's geographic and address lists. The Postal Service decided that the Bureau's proposed cooperative arrangement would be too costly and would not add enough

value to its existing methods for planning mail delivery routes. The Bureau therefore is continuing to explore the availability of information from other sources, such as local governments, for updating its automated geographic system. In particular, the Bureau plans to explore the availability of files and methodologies for performing data exchanges of geographic file information.

#### *Opportunities Provided by a One-Number Census*

We believe a one-number census, which combines the results of the traditional head count supplemented by administrative records and statistical procedures to produce a single, best possible set of numbers by the legal deadlines, offers several advantages. Primarily, a one-number census provides the potential for improving the counts through the use of a combination of methodologies.

In 1990, the Bureau developed two sets of census numbers. The first set resulted from standard census procedures and was produced by the statutory deadlines. The other set was a composite of the first set adjusted by statistical procedures but completed after the Statutory deadlines. In 1990, there was considerable controversy and resulting litigation over which set should be the official census numbers.

Although a one-number census presents operational and technical challenges, using it could improve the count and reduce the overall cost of a census. Obtaining consensus on statistical procedures before the actual census is conducted also could help avoid the controversy that recently surrounded the issue of adjusting the census count. To obtain these benefits, the Bureau must develop operational procedures that will integrate the methodologies selected to produce a one-number census and test them in 1995. This effort will be difficult given the limited time left to prepare for the 1995 test census.

#### NEED TO SCRUTINIZE FEDERAL DATA REQUIREMENTS

In December 1992, OMB asked federal agencies to identify their data needs from the 2000 census. In its request, OMB also specifically asked the agencies to identify possible alternative data sources and to comment on two design alternatives under consideration by the Bureau that would affect data content: expanded content and continuous measurement.

The Bureau is currently reviewing the federal agencies' responses. Preliminary information shows that the agencies want the same type and amount of data collected in the 1990 census. In general, federal agencies did not identify alternative sources to satisfy their data needs. Also, many had reservations about other design alternatives that featured changes in data availability or reliability.

The Bureau and OMB officials need to scrutinize these responses to determine what the Bureau must do to meet federal data needs. In a 1986 review, we found, for example, that obtaining housing data from all households to produce block level data appeared questionable.<sup>4</sup> We also found that user requirements for decennial data were not fully substantiated and evaluated before placing the question on the census form. The challenge is to weigh the benefits of these data against the cost of collecting them.

The fundamental issue underlying the census redesign effort is whether the census can and should continue to be the only source for much of the nation's population and housing information. The National Academy of Science panel, whose chairman is testifying here today, will assist in this systematic review of population and housing data needs. The Academy's panel will be examining the need for data at different frequencies and geographic levels of detail. It also will be examining the availability and costs of alternative methods to obtain these data. We urge officials in the executive branch to examine carefully how to meet federal data needs for subnational data—by a once-in-a-decade census, by more frequent sample surveys that produce less geographically detailed data, or by administrative records.

#### *Expanded Content Through Matrix Sampling Raises Concerns*

Expanded content through matrix sampling would provide data on a wider range of subjects than those in 1990. Under this alternative, which the Bureau is considering as part of its test in 1995, most households would receive a short form. Other households would receive one version of several different medium length forms, each with a different set of questions. (In 1990, a single longer form was sent to approximately one in six households.) With such matrix sampling, more questions could be asked overall without increasing the total respondent burden and the maxi-

<sup>4</sup> See *Decennial Census: Issues Related to Questionnaire Development* (GAO/GGD-86-74BR, May 5, 1986).

mum burden on any one sample household receiving the medium length form would be less than it was in the prior census.

Because matrix sampling presents difficult operational and processing problems, there are limitations, however, to this design. Only subjects collected on the same sample form could be cross-tabulated directly. For example, if one sample form collected information on veteran status and another sample form collected information on income, the Bureau could not produce a cross-tabulation showing the income status of veterans directly from the data. Such information would have to be produced through estimates relying on other data or assumed mathematical relationships. Moreover, to reduce total respondent burden, any one item generally would be asked of fewer households. Thus the data would have less reliability (higher sample error) than it did in 1990. Many data users had reservations about the ability to obtain the cross-tabulations they needed and the reliability of the data at smaller geographic levels.

#### *Continuous Measurement Alternative Depends on Federal Data Needs and Cost*

Under the continuous measurement design, the Bureau would produce basic counts in the census year but would collect various characteristics data on a sample basis throughout the decade. This approach has both positive and negative implications. This approach trades off census sample data for small areas to produce more timely data over the decade for larger geographic areas over the decade. The frequency of the data would vary from 1 year for states and large metropolitan areas to 5 years for smaller areas (below 50,000 population). Many users expressed reservations about the ability of this design to satisfy their needs for data tabulated at lower levels of geography such as census tracts, neighborhoods and small towns, cities, and counties. The Bureau continues to believe that continuous measurement has merit and is conducting further research and evaluation efforts in this area.

The interest in continuous measurement grows out of the desire to obtain more sociodemographic data in between decennial censuses for subnational areas.<sup>5</sup> For example, legislation was passed in 1976 mandating a mid-decade census. The Congress hoped that the cost of conducting such a census would generally be offset by reductions in other statistical series. Because the Bureau could not identify substantial cost offsets, OMB did not request and the Congress did not provide funds to take a mid-decade census.

We believe, given the likely costs of expanding the collection of data between censuses, the only feasible offsets of sufficient magnitude are reductions in the cost of the decennial census. Cost efficiencies appear most probable for the 2000 census by using procedures to obtain greater public cooperation and to use sampling during the follow-up efforts. Moreover, the availability of continually updated address files and geographic data would enhance the accuracy and reduce the cost of intercensal data.

#### **1994 BUDGET REQUEST MAY UNDERFUND SOME IMPORTANT ACTIVITIES**

Based on our review of the Bureau's 1994 budget request, data processing requirements initiative, and discussions with Bureau officials, we are concerned that the fiscal year 1994 budget request has not allowed sufficient funds for some important research and testing activities. The adequacy of census planning in the next few years will have significant implications for the 2000 census and beyond, and the availability of sufficient funding is important for that planning. Prudent investments today can lead to savings tomorrow. Specifically, we question whether the Bureau's 1994 budget request contains sufficient funding for data capture research,<sup>6</sup> improvements in the Bureau's address list and automated geographic system, and 1995 test preparations.

#### *Data Processing Research and Testing May Be Underfunded*

The Bureau has made several assumptions for 2000 census data processing: the primary method for collecting data will be mailout/mailback with respondent friendly rather than processing-friendly questionnaires; the 2000 census workload will be larger than it was in 1990; and the processing schedule will be the same or compressed. The Bureau also seeks to reduce the error level in processing. To satisfy these needs, the Bureau has established a goal of developing a processing system that uses a one-step image scanning device to replace the two-step process that has

<sup>5</sup> See *Federal Data Collection: Measuring Race and Ethnicity is Complex and Controversial* (GAO/T-GGD-93-21, Apr. 14, 1993) and *Federal Formula Programs: Outdated Population Data Used to Allocate Most Funds* (GAO/HRD-90-145, Sept. 27, 1990).

<sup>6</sup> Data capture involves the reading and interpretation of data from census questionnaires.



been used in the past several censuses. The system also incorporates optical character and mark recognition, whereas the existing system only recognizes marks. Because of the risks and uncertainties of developing such a system, the Bureau decided to develop the new system, while maintaining and upgrading its existing one.

The Bureau has prepared a requirements initiative for data capture equipment and engineering support for the year 2000 research and development program. The initiative calls for much of the work to be done through contractor support. The initiative identifies an implementation schedule for completing the tasks but recognizes the uncertainties of the funding. The Bureau requirements initiative identifies a shortfall of about \$1.5 million (of \$3.9 million needed).

Major census automation activities require careful planning, time for testing, and front-end investments. We are concerned that the scenario developing for the 2000 census is reminiscent of the Bureau's automation initiatives planned for the 1990 census. In preparing for the 1990 census, the Bureau eliminated a possible data capture methodology early in its planning phase. We believe the Bureau's decision was influenced by its late start in detailed planning, reluctance to revise the questionnaire form, and a slow procurement process.<sup>7</sup>

In planning for the 2000 census, we thought the Bureau was on a better schedule, but we are now concerned that a lack of funding could eliminate another opportunity to thoroughly explore promising technology. In addition to the funding limitations, the requirements initiative has been delayed for about 5 months in part because of procurement process questions identified by the Commerce Inspector General and currently unresolved within the Department of Commerce. Procurement delays in acquiring minicomputers for the 1990 census led to a \$1.1 million payment to bid protestors and delayed software development and testing and address list development activities.<sup>8</sup>

#### *Geographic Support Activities May Be Underfunded*

We are concerned that the 1994 budget request does not provide sufficient funds for geographic support activities that require long lead times. The budget does not provide for street canvassing to reconcile differences that will inevitably be identified in an imminent test comparing address lists of the Bureau and the Postal Service. Such a reconciliation is needed to determine the accuracy of Postal Service updates for the Bureau's address list. Moreover, the methodology to update the Bureau's automated geographic files is not known at this time. This methodology also may result in the need for additional funds.

#### *Test Census Preparations May Be Underfunded*

We question whether the fiscal year 1994 budget provides sufficiently for preparations necessary for the 1995 test census. Although plans are not complete, the Bureau now is considering four locations for its test, compared to two locations it assumed would be needed when it prepared the 1994 budget. Although this difference may not appreciably affect the resources needed for most planning activities, it could affect the resources needed for updating the geographic information and creating the address list for the four selected sites. These activities need to occur mostly in fiscal year 1994.

#### **LITTLE TIME LEFT FOR FUNDAMENTAL REFORM TO SUCCEED**

The Bureau's efforts to assess design alternatives have consumed valuable time and scarce resources. In the time remaining before the September 1993 deadline, the Bureau, the Department of Commerce, and OMB must quickly focus on those features viable for the 2000 census that offer genuine opportunities to improve the coverage accuracy of the census at a reasonable cost while meeting federal data needs. We also encourage the continued exploration of new ways to collect data between decennial censuses.

As our society has changed in fundamental ways, so must fundamental changes occur in the census. Such census breakthroughs can occur only if they are properly planned, tested, and implemented. Bureau, Commerce, and OMB officials collectively and cooperatively must assume the responsibility for leading this effort. They must clarify the necessity and purpose of fundamental change, establish the proper sense of urgency, develop plans for its implementation, and devote appropriate resources to its accomplishment.

<sup>7</sup> See *Decennial Census: Status of Plans to Computerize Questionnaire Data* (GAO/GGD-86-76BR, May 5, 1986).

<sup>8</sup> See *Decennial Census-Minicomputer Procurement Delays and Bid Protests: Effects on the 1990 Census* (GAO/GGD-88-70, June 16, 1988).



This concludes my prepared statement. My colleagues and I would be pleased to answer any questions.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY CONGRESSMAN SAWYER TO WILLIAM M. HUNT

*Question.* Could you elaborate on the procurement process concerns that are delaying the Bureau's data processing research?

*Answer.* The Bureau identified an opportunity to expedite its data capture research by using the Tennessee Valley Authority's Technical Brokering Program. By using this program, the Bureau would not have been required to pursue the usual, often lengthy, government acquisition process. The Commerce Department's Inspector General's questioned whether the Bureau had the authority to participate in the TVA's program and whether the Bureau was required to use a procurement contract for research and development of scanning device technology. Also, the Inspector General questioned the efficiency and economy of using the TVA program.

The Department's Office of General Counsel determined that the Bureau was not required to obtain the proposed services through a procurement contract, and could use the TVA program. The Department's counsel based its decision on the understanding that the primary purpose of the planned research was the promotion of commercial development of scanning device for the Bureau's direct benefit and use.

According to information available to us shortly after the hearing, the Department had ostensibly resolved the issue by deciding that the Bureau should employ a competitive cooperative agreement itself, rather than using the TVA's program. As of July 16, however, we understand that the Department has determined that a more appropriate vehicle is a "joint" agreement with another party which can be either for profit or not for profit organization. Under a joint agreement there must be a mutuality of interest between the Bureau and the other party.

In spite of these recent developments, we remain concerned about the data capture research activities of the Bureau because,

(1) the time it is taking to resolve this issue could jeopardize the Bureau's research schedule;

(2) a joint agreement requires specific criteria to be followed and, to our knowledge, the Bureau has limited experience with this type of agreement, particularly if a for profit organization is the other party; and

(3) the availability of funding in fiscal year 1994 for this research appears questionable.

*Question.* What are the implications of this delay for the 1995 test?

*Answer.* We believe the delay in determining the method to obtain outside assistance, the additional time needed to follow proper procedures, and the limited availability of funding in fiscal year 1994, rule out the possibility of having a new data capture technology available for testing in the 1995 test.

Mr. SAWYER. Thank you.

Let me ask you the basic question: Is the Bureau's planning program consistent with the Academy's recommendations to this point, projected directions for further investigation and your own observations?

Mr. HUNT. I would say generally it is consistent. I think certainly in the area of sampling for nonresponse it is right on target. I think both have interest in the issue of looking at or analyzing 1990 PES data to get a better fix on gross errors at the block level.

But I think, as Dr. Scarr mentioned, there are some small variations and differences. I think they are going to get together and try to work those out. I think largely, too, they are all interested in administrative records. So while there are shades of difference here and there, I don't see any fundamental problems.

Mr. SAWYER. Are there ways that the Academy's work from here forward can help guide the Bureau's efforts?

Mr. HUNT. Yes, I think probably the most important thing, if I were making the recommendation, would be for the Academy to really focus on assessing Federal data needs. That is going to be

very difficult, and I think it is going to be very controversial. It is not going to be done easily.

I think that the Academy weighing in, coming to the table with its credibility and its independence, would be a major factor in helping policymakers and stakeholders sort through those tough decisions. I think it would also be a brilliant opportunity to identify alternative means of collecting those data.

But I think the most important thing, though, in order to be helpful, is that this contribution really must be on time. I mean, the Academy is going to have to make that November 1994 deadline to make a contribution.

Mr. SAWYER. You give great emphasis to that sense of urgency that is needed. How can we enhance that sense of urgency? What role can OMB play in that process?

Mr. HUNT. Well, I am taking that "we" very broadly. I think, first of all, I would like to say this committee has had a very rigorous oversight schedule which I think has been really very helpful. I think it has helped focus attention on this key issue that I don't think otherwise would have been given.

And I think, also, the staff of this subcommittee have done an excellent job, on a bipartisan basis, I might add, which I think should be held up as a model—

Mr. SAWYER. I would agree.

Mr. HUNT [continuing]. For other people to look at. It has been very effective. I think getting the Bureau and the Postal Service back together was a good example of that. I think another thing, in terms of what the committee might be able to do, I think I mentioned this back in September 1990 when we were getting ready for that census, would be to start thinking about touching base with the administration to see where are we with getting a Bureau Director in there. We need someone nominated formally and confirmed, as well as I might add, Mr. Chairman, other folks at the Commerce Department. We need key officials who are going to make some decisions here, because there is no substitute for leadership. I think that is what is really, really much needed here. And I think OMB plays a role.

You mentioned OMB, I don't want to neglect it. I think it is obviously very important. It is key here in terms of figuring out what the Federal data needs might be. I think it is also important to keep the Bureau focused to make sure they will make breakthroughs. I think it is important in terms of making sure there are adequate resources available for the investments, if you will, the initial investments needed to make change in the future. I think that is key.

And I might commend OMB officials, I think they have done an excellent job. In years past, they have sometimes doubled what the Department and what the Bureau had asked for in terms of this area, because they see the value of it. I would say that is important.

The last thing I would say is that there are a lot of short time frames here. A lot of clear answers are going to be needed; forms are going to be needed to be cleared through the Commerce and the OMB processes. I think everybody has got to pull together to

make sure those things get done in a timely way so we don't have any unusual delays.

Mr. SAWYER. Well, I try not to miss an opportunity to at least touch base on that sort of thing. As I mentioned to Dr. Scarr, the only job that may be tougher than coming in being the new director is continue to try and manage all of this from the position of an acting director.

Mr. HUNT. Yes.

Mr. SAWYER. That is something that we touch base with on a regular basis.

Mr. HUNT. An awful lot of what is needed is not just internally focused, but having a voice externally. Playing that relational role is very important.

Mr. SAWYER. It is not the sort of thing that a committee of Congress can do very well.

Mr. HUNT. You can't do it on a day-to-day basis and that is the difference.

Mr. SAWYER. That is right.

You talked about potential for fundamental breakthroughs. I mean, I assume that your definition of fundamental breakthrough is narrowing. What did you have in mind?

Mr. HUNT. Well, I think what we are talking about in our statement, is that we have always had the view that the census is not going to be reformed in one census. The time is simply not there. But I don't think we can ever give up hope.

I think we need to keep the budget situation in mind, and we also know the data qualities have been deteriorating over the last several censuses. I think the data show that the longer the census goes on through all these subsequent coverage improvement efforts, that the data quality really does deteriorate. It seems to me you get in and get out quickly.

Mr. SAWYER. You mean the actual duration?

Mr. HUNT. Yes, the duration of the effort. I think census errors were about 2.8 percent in the early part of the census, and it jumps up 10 times that by the latter part of the census. And you are spending, as you know, a lot more money at the end of the process. A lot of money is going in and you are getting a lot worse data.

So I think you need to keep these fundamentals in mind. But I think if we stick to what is realistic for 2000, that would include sampling for nonresponse, a better coordination with the Postal Service on the address lists, improving and streamlining the census questionnaire to improve the response rate. Making it more user-friendly I think would be important, as well as possibly pursuing the objective of a one number Census. Those are some things I have in mind.

Mr. SAWYER. Have the Federal departments been helpful in finding alternative data source selections?

Mr. HUNT. Well, as I think I said in my statement, I don't think fundamentally they have. They basically have not raised any alternative that already isn't in use. If you don't mind, Mr. Chairman, I would like to ask Mr. Kaufman, who has been examining some of those data to respond. He can give you a more complete answer to that.

Mr. KAUFMAN. Well, generally, the Federal data users in response to the OMB letter have commented that there are no alternative data sources that provide the required level of geographic data and the cross tabulations that they are looking for.

For example, the Veterans Administration does its own national data survey, but it is on about 10,000 veterans, so it doesn't provide useful data at lower geographic levels. The VA also sponsors questions on the Current Population Survey, but here again, that is basically a national survey which can only provide national data and data for States and major metropolitan areas. So here again, it is not at the level that they would like.

Mr. SAWYER. You mentioned in your testimony that something like a half a billion dollars could have been saved if the Bureau had conducted a 10 percent nonresponse sample. That is a substantial savings and not something to be sniffed at.

If the 10 percent sample were not sufficient, how does the size of that sample relate to the potential for savings? Just so we could have a perspective.

Mr. HUNT. Well, obviously the larger the sample, the savings diminish as a result, but they are not insubstantial. For example, I think our estimate is a 10-percent sample would have saved about \$460 million in 1990. If we move that 10 percent up to 20 percent, I think we then estimate the savings would be about \$425 million. If we go up to 33 percent, I think the savings are somewhere in the vicinity of about \$325 million. And if you go up to a 50 percent sample, savings are something around \$215 million. So there are savings there, at any of those levels.

Mr. SAWYER. I am assuming that when you cite those higher percentage samples, that those really are on the far end of what might be deemed necessary, or have you not taken a look at that?

Mr. HUNT. We haven't ourselves taken a look at it, but maybe, Jack, do you have a response?

Mr. SAWYER. What do others say? What level of sampling would be necessary to provide an appropriate level of comfort?

Mr. KAUFMAN. I mean, the big issue which will have to be resolved by some statistical analysis, is at what point a sample provides lower quality data than some of the costly follow-up activities. It has often been said that when you get to the latter stages of follow-up, you are getting surrogate and last resort information. So I think the Bureau is trying to do an analysis to determine where the break-even point is.

Mr. HUNT. I might add also, Mr. Chairman, as I mentioned before, the further down the Census road you go, the greater the cost and the less accurate the data. So it all has to fit into this equation. So, actually, it is very possible that a sample would probably save a lot of money. As I said, 10 percent would save \$460 million.

I can't sit here this afternoon and swear that is the number we ought to be shooting for right now. The research I think would give us the information we need to help make that judgment. But I think there are savings through sampling.

Mr. SAWYER. You mentioned the cooperative work that has gone on with the Postal Service, address list sharing and so forth. Are



there—we have done some work in trying to get a sense of barriers that may exist with regard to one way sharing of address lists.

Are there upstream problems that we ought to be aware of?

Mr. HUNT. I think we are all encouraged by the renewed efforts between the Bureau and the Postal Service to work out an arrangement between them. However, I think there is still a bit of a problem and some uncertainty. I think the Postal Service still perceives that they have a problem in providing the precise addresses.

In other words, the situation they conceive is that the Bureau would provide its addresses to the Postal Service, the Postal Service would review them and would identify, if you will, the magnitude of differences. In other words, they may find that a block comes in and the Bureau says there are a hundred addresses, and the Postal Service counts 125, for example. So there is obviously a difference of 25.

The Postal Service right now feels they are constrained in providing back to the Bureau the 25 specific addresses that are on their list that aren't on the Bureau's list. They are willing, apparently, to provide the gross difference, but that provides complications for the Bureau because they have to go out and try to figure out where are those addresses and whether the Postal Service in its records has their version of those addresses elsewhere.

I think there is a bit of a serious problem from our point of view, Mr. Chairman. We have not looked at this ourselves in any depth, but I do know that in 1990 census the Postal Service, in its various reviews of address lists, provided millions of specific addresses back to the Bureau during that period.

But I guess the real answer here for you, Mr. Chairman, is if there are any differences or any concerns here, and I know there have been some statutory issues, they can be readily resolved with legislation. I would encourage the Bureau and the Postal Service to come to some conclusion fairly quickly on that, and sit down and discuss it with you and others as needed.

Mr. SAWYER. I am driven by that perceived difficulty, in any case, to ask if there is trouble with something as fundamental as address lists. What implication does that have for more complex interaction with a wider range of administrative records and are those as overcomeable as the circumstances you just described?

Mr. HUNT. Well, I think that is why we concluded that while administrative records are very promising over the long run, I don't think they are going to be something that realistically is going to be available for 2000 in any level. So I think that you are right. I mean, there are all sorts of issues in terms of the level of detail and the accuracy of the data that are in those administrative records.

Also, there is a huge intergovernmental dimension to it, and I have always been one who thought that we need to look at this issue from an intergovernmental perspective over the long run. I don't mean just the decennial census, I mean the entire statistical system itself. I think there is a lot of duplication and overlap, et cetera, in that system, and I think we can get a better return on our collective, if you will, public investments in the long run if we think of it in those global terms. But that is a few leagues away from where we are now. But undoubtedly there would be problems.

Mr. SAWYER. Well, that notion of long-term, collective public investment in this really is what drives the questions that we are trying to ask about getting beyond this illusion of precision when timeliness is critical in a period of change. But continuity, over a longer period of time, is just as important.

I am not sure that I am going to frame this as a question, but it is certainly a concern that this subcommittee has, in assuring that in a time of rapid change, we get information that is useful and relevant without undermining the long-term continuity of data that scholars and others really rely on.

Mr. HUNT. Well, we would agree with that too, Mr. Chairman. I mean, obviously we are very fond of comparability and continuity in data. But if it is not relevant, it is not worth much.

Mr. SAWYER. Exactly. We will probably have other questions for you, but for today, I think we have dodged the time bullet.

If there is no further business to come before the committee, I thank all our witnesses, and we stand adjourned.

[Whereupon, at 2:34 p.m., the subcommittee was adjourned.]





BOSTON PUBLIC LIBRARY



3 9999 05982 768 1





ISBN 0-16-043630-3

